Minimal Covers of Formal Languages

by

Michael Domaratzki

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Mathematics

in

Computer Science

Waterloo, Ontario, Canada, 2001

I hereby declare that I am the sole author of this thesis.

I authorize the University of Waterloo to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the University of Waterloo to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

Let $L_1, L_2 \subseteq \Sigma^*$ be languages. If $L_1 \subseteq L_2$, then we say that L_2 covers L_1. Let $\mathcal{C}, \mathcal{D}$ be classes of languages over Σ. We call a cover $L_2 \in \mathcal{C}$ of L_1 a $\mathcal{C}$-minimal cover with respect to $\mathcal{D}$ if, for all $L' \in \mathcal{C}$, $L_1 \subseteq L' \subseteq L_2$ implies $L_2 - L' \in \mathcal{D}$.

We examine in detail the particular cases of minimal regular covers and minimal context-free covers, both with respect to finite languages. The closure properties of the class of languages which possess a minimal cover (regular or context-free) are examined. We provide a complete characterization of bounded languages which have a minimal regular cover. An analogous results for languages bounded by $w_1^* w_2^*$ and minimal context-free covers is also given.

The problem of whether a given (linear) CFL has a minimal regular cover is shown to be undecidable, while it takes exponential time infinitely often on a Turing machine to decide whether a given bounded CFL has a minimal regular cover.

Acknowledgements

C. Pennell and E. Domaratzki: my greatest supporters and my first teachers.

J. Shallit: my supervisor who challenges me but refuses to let me beat myself up.

M. Mishna: encouragement and proof reading of this thesis.

crew: shouts out to kristy, rob, reuben and kerri. bling bling.

Contents

Chapter 1

Introduction

Approximation of languages by regular languages is an important problem in systems with limited resources. For example, the problem for approximating finite languages by infinite regular languages has received attention recently [10, 41]. We are interested in approximating non-regular infinite languages by regular languages. To this end, we consider regular languages which are minimal covers of non-regular languages. We extend the notion of simplicity and immunity from structural complexity [13, 11, 42] to sub-recursive classes of languages, specifically regular languages, to introduce the notion of minimal regular covers.

For languages L_1, L_2 over a common alphabet Σ, we say that L_2 *covers* L_1 if $L_1 \subseteq L_2$. In this thesis, we are concerned with languages which possess covers that lie in a given complexity class, and that are minimal, in some sense, over all covers from that complexity class. In particular, we propose the following definition:

Definition 1.0.1 *Let $\mathcal{C}, \mathcal{D}$ be a classes of languages over an alphabet Σ^*, and let $L_1 \subseteq L_2 \subseteq \Sigma^*$. Then L_2 is a minimal $\mathcal{C}$-cover of L_1 with respect to $\mathcal{D}$ if $L_2 \in \mathcal{C}$*

and for all $L \in \mathcal{C}$ *such that* $L_1 \subseteq L \subseteq L_2$*, we have* $L_2 - L \in \mathcal{D}$*.*

This definition is the basis for the work which follows in Chapters 3 and 4.

If $\mathcal{D}$ is the class of finite languages and $\mathcal{C}$ is a class of languages which is a boolean algebra under the operations of union and complementation, we can show that minimal $\mathcal{C}$-covers with respect to finite languages are unique, up to finite modification. This theorem is an easy generalization of Theorem 3.0.1.

We will be concerned primarily with the case where $\mathcal{C}$ is the class of regular languages and $\mathcal{D}$ is the class of finite languages. In this case, we simply say that L_1 has *minimal regular cover* L_2. Theorem 3.6.5 demonstrates that not every language has a minimal regular cover, even among the context-free languages.

1.1 Preliminaries

We first establish some preliminaries of formal language theory which will be used in this thesis. For an additional reference about formal language theory, see Hopcroft and Ullman [24]. For preliminaries of complexity theory, see Papadimitriou [39] or Balcázar *et al.* [2]. Any notion not defined in this section can be found in Hopcroft and Ullman [24], including the formal models of Turing machines, finite automata, context-free grammars and context-sensitive grammars.

Let Σ be a finite alphabet and let Σ^* be the set of finite words over Σ. A language L over Σ^* is a subset of Σ^*. For strings $x \in \Sigma^*$, $|x|$ denotes the length of x. For any letter $a \in \Sigma$, $|x|_a$ is the number of occurrences of a in the word x. For any integer $m \geq 0$, let $\Sigma^{\leq m} = \{x \in \Sigma^* \ : \ |x| \leq m\}$.

For any string $x \in \Sigma^*$, $x = x_1 x_2 \cdots x_n$, where $x_1, x_2, \ldots, x_n \in \Sigma$, the reversal of x, denoted by x^R, is defined to be $x^R = x_n x_{n-1} \cdots x_2 x_1$.

For a given alphabet Σ and a language $L \subseteq \Sigma^*$, we let $\overline{L}$ denote the complement of L, that is, $\overline{L} = \Sigma^* - L$. For any finite language L, $|L|$ denotes the cardinality of L.

We denote the class of languages recognizable in polynomial time by a deterministic Turing machine by **P**, and the class of languages recognizable in polynomial time by a non-deterministic Turing machine by **NP**. The class of languages recognizable in linear exponential time on a deterministic Turing machine is denoted **E**. Other classes of languages are given by upper-case script letters: $\mathcal{C}, \mathcal{D}$, etc.

We denote the class of regular languages, that is, those recognizable by a finite automaton, by **REG** while the classes of finite languages, context-free and context-sensitive languages are denoted by **FIN**,**CF** and **CS**, respectively. The class of deterministic context-free languages (DCFLs) are denoted by **DCF**.

A language is said to be recursive if it is accepted by a Turing machine that halts on all inputs. A language that is accepted by a Turing machine (which does not necessarily halt on all inputs) is said to be recursively enumerable (r.e.). The classes of recursive and recursively enumerable languages are denoted by **REC** and **RE**, respectively.

We have the following hierarchy of strict inclusions (see Hopcroft and Ullman [24]):

$$\mathbf{FIN} \subsetneq \mathbf{REG} \subsetneq \mathbf{CF} \subsetneq \mathbf{CS} \subsetneq \mathbf{REC} \subsetneq \mathbf{RE}. \tag{1.1.1}$$

For any class of languages $\mathcal{C}$, let co-$\mathcal{C}$ denote the class of languages whose complement is in $\mathcal{C}$. That is, co-$\mathcal{C} = \{L \ : \ \overline{L} \in \mathcal{C}\}$.

For any function $f : X \to Y$ and any subset $Z \subseteq X$, the restriction of f to Z, denoted by $f|_Z : Z \to Y$, is the function defined by $f|_Z(z) = f(z)$ for all $z \in Z$.

1.2 Synopsis

In Chapter 2, we will examine concepts in computational complexity theory which are the analogues of minimal covers for classes such as **P**, **NP** or **RE**. In particular, Balcázar and Schöning [3] note that Post [42] introduced the concept of simple sets to recursive function theory in an attempt to define languages which were neither r.e.-complete nor recursive.

The definitions of immune and simple have been used since Post's introduction by many authors, including Flajolet and Steyaert [13], who transformed Post's notion of immunity in recursive function theory to the notion of immunity pertaining to complexity classes. Immunity and simplicity has also been extended by Balcázar and Schöning [3], who studied bi-immune sets.

While the concepts of immunity and simplicity have a strong standing in the literature, they are focused primarily at classes situated near the highest level of the Chomsky hierarchy. However, Chapter 2 will also examine work similar to minimal covers pertaining to the context-free languages. In particular, the work of Hunt *et al.* [28, 27], while concentrating primarily on the difficulty of problems concerning the similarity of regular expressions and CFGs, contains results which characterize a class of CFLs which are **REG**-immune.

We also have the concept of minimality pertaining to regular languages developed by Câmpeanu *et al.* [10], where the authors define the concept of a cover-automaton. These cover-automata are defined in order to give succinct representations of finite languages by infinite regular languages.

In Chapters 3 and 4, we develop the notions of minimal covers of formal languages. Chapter 3 examines minimal regular covers in depth. The closure properties of the class of languages possessing a minimal regular cover are examined in Sections 3.1 and 3.2. We give a new proof of a theorem of Ginsburg and Spanier [18] on bounded regular languages, and use this theorem to characterize completely those bounded languages with a minimal regular cover. We also examine several decision problems relating to minimal regular covers. Chapter 4 considers results on minimal context-free and minimal co-context-free covers. We conclude with Chapter 5, in which we give several open problems relating to minimal covers.

Chapter 2

Related Work

This chapter examines the existing work related to minimal covers. We begin with the theory of immunity and simplicity related to sets of integers introduced by Post [42] in 1944. We continue with the natural extension of immunity to classes of formal languages, through the paper of Flajolet and Steyaert [13], and examine some results on immunity and bi-immunity.

2.1 Immunity and Recursive Function Theory

Post [42] initiated research on simple sets in recursive function theory, where he defined them as follows:

Definition 2.1.1 *A r.e. set A is* **simple** *if $\overline{A}$ is infinite and contains no infinite r.e. subset.*

Post then proved the existence of a simple set, as well as the following lemma:

Lemma 2.1.2 *Every infinite recursively enumerable set contains an infinite recursive subset.*

Thus there are no languages in co-**RE** − **REC** which are simple.

Dekker [11], improving directly on Post's results, gave the following definition of immunity:

Definition 2.1.3 *A set A is* **immune** *if A is infinite and A has no infinite r.e. subset.*

The main result of Dekker [11] gives some closure properties of simple (r.e.) sets:

Theorem 2.1.4 *The following hold for simple sets:*

- *The product of two simple sets is simple.*
- *The sum of two simple sets is either simple or is co-finite.*
- *There exist two simple sets whose union equals $\mathbb{N}$.*

Post introduced simple sets in an attempt to determine whether there exist **RE** sets which are neither **RE** complete nor recursive, though Balcázar and Schöning [3] note that simplicity failed as a method for solving Post's problem. However, the notions of simplicity and immunity, which we see below were quickly adapted to a computational complexity-theoretic notion, have remained important concepts in the literature.

2.2 Immunity and Complexity Theory

Flajolet and Steyaert [13] extended the definitions of immunity and simplicity to move them from a recursive function theory setting to a computational complexity theory setting. This entails defining immune and simple languages for every language class $\mathcal{C}$.

Definition 2.2.1 *Let $\mathcal{C}$ be a class of languages.*

- *A language L is $\mathcal{C}$-immune if L is infinite and no infinite subset of L is a member of $\mathcal{C}$.*
- *A language is L is $\mathcal{C}$-simple if L is co-infinite and intersects every infinite member of $\mathcal{C}$.*

It is an easy observation that L is $\mathcal{C}$-simple if $L \in \mathcal{C}$ and $\overline{L}$ is $\mathcal{C}$-immune. Given this definition, we have the following lemma:

Lemma 2.2.2 *Let $\mathcal{C}$ be a class of languages closed under complement. A co-infinite language $L \subseteq \Sigma^*$ is $\mathcal{C}$-simple if and only if it has minimal $\mathcal{C}$-cover Σ^*.*

Proof:

$$
\begin{aligned}
L \text{ is } \mathcal{C}\text{-simple} &\iff \overline{L} \text{ has no infinite } \mathcal{C} \text{ subset} \\
&\iff (\forall S)(\overline{S} \subseteq \overline{L} \text{ and } \overline{S} \in \mathcal{C} \Rightarrow \overline{S} \text{ is finite }) \\
&\iff (\forall S)(L \subseteq S \subseteq \Sigma^* \text{ and } S \in \mathcal{C} \Rightarrow \Sigma^* - S \text{ is finite }) \\
&\iff L \text{ has minimal } \mathcal{C}\text{-cover } \Sigma^*.
\end{aligned}
$$

□

Flajolet and Steyaert show that every denumerable class of subsets of $\mathbb{N}$ contains a simple set. Their standard diagonalization technique is an adaptation of Post's original construction of a simple set. Thus for each class of languages, there is a simple language for that class. However, we do not know in which class the simple language lies.

Of the results given by Flajolet and Steyaert [13], we are most concerned with those contained in Flajolet and Steyaert's section entitled "Application to subrecursive classes", since we will be considering minimal covers from subrecursive classes in Chapters 3 and 4. Flajolet and Steyaert note that the well-known **CF** language $\{a^n b^n \ : \ n \geq 1\}$ is a regular-immune language, while the **CS** language $\{a^n b^n c^n \ : \ n \geq 1\}$ is a **CF**-immune language. Both of these results can be obtained by simple applications of pumping-type lemmas.

An immediate application of immunity to complexity theory is the following easy observation:

Lemma 2.2.3 *If there exists an* $\mathbf{P}$*-immune set* $A \in \mathbf{NP}$ *then* $\mathbf{P} \neq \mathbf{NP}$*.*

Proof: If A is $\mathbf{P}$-immune and $A \in \mathbf{NP}$ then certainly $A \notin \mathbf{P}$ since A has no infinite $\mathbf{P}$-subset, including A itself. □

As another example of results in immunity theory, we present the following nice result of Blum [7, 8] which we can translate in terms of immunity:

Theorem 2.2.4 *Let* M_i *be an enumeration of Turing machines. Then the following language* $L_{\min}$ *is* **REC***-immune:*

$$L_{\min} = \{M_i \ : \ \forall j < i, L(M_i) \neq L(M_j)\}$$

where $L(M_i)$ is the language accepted by the Turing machine M_i.

2.2.1 Bi-immunity

Balcázar and Schöning [3] have extended the notion of immunity by defining the concept of bi-immunity:

Definition 2.2.5 *A language L is bi-immune for $\mathcal{C}$, or simply $\mathcal{C}$-bi-immune, if both L and $\overline{L}$ are $\mathcal{C}$-immune.*

This appears to be the first definition of bi-immunity, though it was implicitly studied in papers by Hartmanis and Berman [4] and Ko and Moore [30]. Bi-immunity is related to the notion of complexity cores, introduced by Lynch [36]. For any language A, the set C is a complexity core of A if, for any Turing machine M accepting A and every polynomial p, there are only finitely many strings $x \in C$ whose computation by M takes time at most $p(|x|)$. It is noted by Balcázar and Schöning [3] that $A \subseteq \Sigma^*$ is $\mathbf{P}$-bi-immune if and only if Σ^* is a complexity core for A.

There are many papers considering bi-immunity. We note only some particularly interesting results here. Let $\mathbf{E}$ denote the set of languages which can be recognized in linear exponential time, that is, $\mathbf{E} = \mathbf{DTIME}(2^{O(n)})$. Ko and Moore [30] note that Hartmanis and Berman [4] proved that there exists a $\mathbf{P}$-bi-immune set in $\mathbf{E}$. In fact, Mayordomo [38] proved that the class of $\mathbf{P}$-bi-immune languages has measure 1 in $\mathbf{E}$, under the resource-bounded measure of Lutz [34, 35], which gives a suitable interpretation of the Lebesgue measure for recursive language classes.

2.3 CFLs and Coverings

We begin an examination of work on CFLs and coverings with an examination of the paper of Hunt, Rosenkrantz and Szymanski [28]. This paper was concerned primarily with the complexity of several decision problems relating to grammars and regular expressions, but also characterizes context-free languages, beginning with the following definition:

Definition 2.3.1 *A CFL L_0 is said to be* **decipherable** *if and only if for all strings u, v, w, x, y in Σ^* with $v, x \neq \epsilon$ such that $\{uv^iwx^iy \; : \; i \geq 0\} \subseteq L_0$, we have*

$$uv^iwx^iy = uv^jwx^ky \Rightarrow i = j = k.$$

With this definition, we have the following theorem of Hunt *et al.* [28]:

Theorem 2.3.2 *Let L be any infinite CFL. Then either L has an infinite regular subset or L is decipherable.*

In Section 3.6, we will see an expanded proof of Theorem 2.3.2 found independently of Hunt *et al.* [28]. As we have seen, if a language L has an infinite regular subset, it does not have minimal regular cover Σ^*. Thus, languages which are not decipherable do not have minimal regular cover Σ^*.

A similar result to Theorem 2.3.2 is also given by Latteux and Thierrin [31, Lemma 5.5, p. 14]. We will return to this in Section 3.6.

2.3.1 Grammatical Coverings

Another notion of coverings, that of grammatical coverings, is given by Gray and Harrison [19]. Recall that $h : A^* \to B^*$ is a homomorphism if $h(a_1 a_2) = h(a_1)h(a_2)$ for all $a_1, a_2 \in A^*$.

Definition 2.3.3 *Let $G_1 = (V_1, \Sigma, P_1, S_1)$ and $G_2 = (V_2, \Sigma, P_2, S_2)$ be context-free grammars. Let $h : P_1^* \to P_2^*$ be a homomorphism. Then G_1* **left covers** *G_2 under h if, for all $w \in \Sigma^*$,*

- *whenever $S_1 \Rightarrow_L^{\pi} w$, then $S_2 \Rightarrow_L^{h(\pi)} w$, and*
- *whenever $S_2 \Rightarrow_L^{\pi} w$ then there exists π' such that $h(\pi') = \pi$ and $S_1 \Rightarrow_L^{\pi'} w$.*

The concept of G_1 being a right cover of G_2 is defined symmetrically. Gray and Harrison [19] contrast their definition to one given by Reynolds and Haskell:

Definition 2.3.4 *Let $G_1 = (V_1, \Sigma, P_1, S_1)$ and $G_2 = (V_2, \Sigma, P_2, S_2)$ be context-free grammars. Let $h : P_1^* \to P_2^*$ be a homomorphism. Then G_2 is said to be a* **weak cover** *of G_1 if*

- *$h(S_1) = S_2$, and*
- *if $A \to x$ is a production in P_1 then $h(A) \Rightarrow^* h(x)$ in G_2.*

Note that if G_2 left (or right) covers G_1 under the homomorphism h, then $h(A) \to h(x)$ is a production in P_2 for each production $A \to x$ in P_1. So G_2 is also a weak cover for G_1.

However, these concepts of covering are strictly concerned with grammars, and not with language coverings as sets. In fact, if G_1 left covers G_2 then $L(G_1) =$

$L(G_2)$ [28]. Hunt *et al.* note that the concept of grammatical covering are not as general as language equivalence. That is, there exist **CF** grammars G_1, G_2 such that $L(G_1) = L(G_2)$ but G_1 does not cover G_2 or vice-versa [28].

Unfortunately, the work of Hunt *et al.* does not consider the relationships between grammars which satisfy $L(G_1) \subseteq L(G_2)$. The concepts of left and right covers are only developed by Hunt *et al.* as a means of classifying accompanying decision problems, eg. $\{(G_1, G_2) \ : \ G_1$ left covers $G_2\}$ is not **RE** for arbitrary CFGs G_1 and G_2 over a binary alphabet, but is **PSPACE**-complete if G_1 and G_2 are linear CFGs [28].

2.3.2 Deterministic CFLs

We also have a brief note of a result by Hunt and Rosenkrantz [27], which we can interpret in terms of bi-immune sets. Recall that a language L is bounded if $L \subseteq w_1^* \cdots w_n^*$ for some words $w_1, \ldots, w_n$. If there do not exist any such words, L is said to be unbounded.

Theorem 2.3.5 *For any DCFL $L \subseteq \Sigma^*$ with $|\Sigma| \geq 2$, either L or $\overline{L}$ contains an regular subset R which is unbounded.*

Upon noting that any DCFL over a unary alphabet is regular, we can rewrite Theorem 2.3.5, using the terminology of Balcázar and Schöning, as follows:

Theorem 2.3.6 *There are no* **REG**-*bi-immune deterministic CFLs.*

2.4 Density Problems

The question of containment of languages is related to a problem of Bucher [9] which remains open: given context-free languages L_1 and L_3 with $L_1 \subseteq L_3$ and $L_3 - L_1$ infinite, must there exist a context-free language L_2 such that $L_1 \subseteq L_2 \subseteq L_3$ with $L_3 - L_2$ and $L_2 - L_1$ both infinite?

This question is true for regular languages, as is noted by Bucher [9], since the regular languages are closed under complement. To see this, note that in the equivalent formulation for regular languages, $L_3 - L_1$ is an infinite regular language. It is an easy task then to construct a regular language L_4 which is an infinite subset of $L_3 - L_1$ and for which $(L_3 - L_1) - L_4$ is also infinite. Thus $L_2 = L_4 \cup L_1$ satisfies the conditions. Similar methods also exist to solve the problem positively for deterministic context-free and recursive languages. However, the lack of closure under complement means these methods do not apply to the CFLs. Thus, the problem of Bucher still remains open.

2.5 Minimal Cover-Automata

The concept of cover-automata was developed by Câmpeanu *et al.* [10] for giving compact representations of finite languages as finite automata.

Given a finite language L, where the longest string in L is of length ℓ, a cover-automaton of L is a deterministic finite automaton $M = (Q, \Sigma, \delta, q_0, F)$ such that $L(M) \cap \Sigma^{\leq \ell} = L$. A minimal cover-automaton of L is a cover-automaton $M = (Q, \Sigma, \delta, q_0, F)$ such that $|Q|$ is minimal among all cover-automata for L. Câmpeanu *et al.* [10] demonstrate that a given finite language may have several non-isomorphic minimal automata.

In terms of minimal covers, given any finite language F, then for any class $\mathcal{C}$ with $\mathbf{FIN} \subseteq \mathcal{C}$, L is a $\mathcal{C}$-minimal cover of F with respect to $\mathbf{FIN}$ if and only if $F \subseteq L$ and L is finite. However, in constructing the minimal cover-automaton M for F, the language accepted by M is not necessarily finite.

Given a DFA M with n states accepting a finite language F, Paun *et al.* have also given an $O(n^2)$ algorithm for constructing a minimal cover automata for a finite language [41].

Chapter 3

Minimal Regular Covers

The focus of this chapter is the topic of minimal regular covers with respect to finite languages, or simply, minimal regular covers. Before beginning the closure properties of minimal regular covers, we prove that minimal regular covers are unique up to finite modification.

Theorem 3.0.1 *Let L be a language, and suppose R, R' are minimal regular covers of L. Then $(R - R') \cup (R' - R)$ is finite.*

Proof: Suppose R, R' are minimal regular covers of L. Let $R'' = R \cap R'$. Then R'' is regular and also covers L, and $L \subseteq R'' \subseteq R$. It follows that $R - R''$ is finite. Similarly, $R' - R''$ is finite. Since $(R - R') \cup (R' - R) = (R - R'') \cup (R' - R'')$, the result follows. □

The following result, due to Shallit, shows us that the languages with minimal regular cover Σ^* are dense:

Theorem 3.0.2 *Almost all languages L have minimal regular cover Σ^*.*

Proof: We show that almost all languages have minimal regular cover Σ^*. Let a language L be chosen at random, that is, for all strings $x \in \Sigma^*$, the probability that $x \in L$ is $1/2$.

Assume that there exists a regular language R with $L \subseteq R$ and $\Sigma^* - R$ infinite. Since $\Sigma^* - R$ is an infinite regular language, there exist words u, v, w with $v \neq \epsilon$ such that $uv^iw \in \Sigma^* - R$ for all $i \geq 0$. But since L is chosen at random, with probability 1 at least one of $uv^iw \in L$ for some $i \geq 0$. Thus L is not a subset of R. This is a contradiction. Thus L has minimal regular cover Σ^*. □

Section 3.1 examines several operations which preserve minimal regular covers, while Section 3.2 examines remaining operations which do not preserve minimal regular covers. We then turn specifically to bounded languages in Section 3.3. Using a characterization of bounded regular languages due to Ginsburg and Spanier, we give a theorem which allows us to determine whether a given language has a regular language R as minimal regular cover.

Sections 3.4 and 3.5 then examine questions of decidability and complexity of some problems related to minimal regular covers. In Section 3.6, which considers some interesting classes of languages and their minimal regular covers. We conclude by briefly considering bi-minimal regular covers in Section 3.7.

3.1 Closure Properties of MRCs

3.1.1 Closure under binary operations of MRC

We examine the closure properties of minimal regular covers under several operations.

Theorem 3.1.1 *Let L_1 and L_2 be languages with minimal regular covers R_1 and R_2, respectively. Then $L_1 \cup L_2$ has minimal regular cover $R_1 \cup R_2$.*

Proof: Let R_3 be a regular language such that $L_1 \cup L_2 \subseteq R_3 \subseteq R_1 \cup R_2$. Note that $(R_1 \cup R_2) - R_3$ infinite implies $R_1 - R_3$ infinite or $R_2 - R_3$. Without loss of generality, assume that $R_1 - R_3$ is infinite. But $R_1 - R_3 = R_1 - (R_1 \cap R_3)$, and $L_1 \subseteq (R_1 \cap R_3) \subseteq R_1$. Thus, R_1 is not a minimal regular cover of L_1, a contradiction. □

Theorem 3.1.2 *If L_1 has minimal regular cover R, and $L_2 \subseteq R$, then $L_1 \cup L_2$ has minimal regular cover R.*

Proof: Let R' be a regular language with $L_1 \cup L_2 \subseteq R' \subseteq R$. But then also $L_1 \subseteq R' \subseteq R$. Since R is a minimal regular cover of L_1, this gives that $R - R'$ is finite, and so R is a minimal regular cover of $L_1 \cup L_2$. □

The following, due to Shallit, shows that minimal regular covers are closed under concatenation with a finite language. We will see in Section 3.2 that minimal regular covers are not closed under arbitrary concatenation, even with regular languages.

Theorem 3.1.3 *Let L be a language having minimal regular cover L and let F be a finite language. Then FL (resp., LF) has minimal regular cover FR (resp., RF).*

Proof: We prove the result for FL, leaving the result for LF to the reader.

By Theorem 3.1.1 it suffices to prove the result when F is a single word, say $F = \{w\}$. So suppose, to get a contradiction, that R_1 is a regular language such that $wL \subseteq R_1 \subseteq wR$, with $wR - R_1$ infinite.

Let $R_2 = \{x \ : \ wx \in R_1\}$. It is clear that R_2 is regular, and since $R_1 \subseteq wR$, we have $R_1 = wR_2$. We claim that $L \subseteq R_2 \subseteq R$. For the first inclusion, suppose $y \in L$. Then $wy \in wL \subseteq R_1$. But then $y \in R_2$. For the second, suppose $z \in R_2$. Then $wz \in R_1$. But $R_1 \subseteq wR$, so $wz \in wR$. Hence $z \in R$.

Since R is a regular cover of L, we must have $R - R_2$ is finite. Then $w(R - R_2) = wR - R_1$ is finite. But this contradicts our earlier assumption. It follows that wR is a minimal regular cover of wL. □

The following theorem, due to Shallit, shows that minimal regular covers are closed under intersection with regular languages.

Theorem 3.1.4 *Let L be a language with minimal regular cover R. Let R' be an arbitrary regular language. Then $L \cap R'$ has minimal regular cover $R \cap R'$.*

Proof: Let R'' be a regular language satisfying $L \cap R' \subseteq R'' \subseteq R \cap R'$. Assume, to arrive at a contradiction, that $(R \cap R') - R''$ is infinite. Define

$$R_0 = R - ((R \cap R') - R''),$$

and note that $L \subseteq R_0 \subseteq R$. Also, $R - R_0 = (R \cap R') - R''$ and so is infinite. Thus, R is not a minimal regular cover R. This is our contradiction. □

Next, we examine quotient. Recall that for languages $L_1, L_2 \subseteq \Sigma^*$, the quotient L_1/L_2 is defined by

$$L_1/L_2 = \{x \in \Sigma^* \ : \ \exists y \in L_2 \text{ such that } xy \in L_1\}.$$

The following result, due to Shallit and the author, shows that minimal regular covers are preserved under quotient by arbitrary languages.

Theorem 3.1.5 *Let L_1 be a language with minimal regular cover R_1, and L_2 be an arbitrary language. Then L_1/L_2 has minimal regular cover R_1/L_2.*

Proof: Clearly if $L_1 \subseteq R_1$, then $L_1/L_2 \subseteq R_1/L_2$. Furthermore, it is well known that if R is regular, then R/L is regular for all languages L [24, Thm. 3.6, pp. 62–63]. Hence R_1/L_2 is a regular cover of L_1/L_2.

Suppose, contrary to what we want to prove, that R_1/L_2 is not a minimal regular cover of L_1/L_2. Then there exists a regular language R' such that

$$L_1/L_2 \subseteq R' \subseteq R_1/L_2$$

and $R_1/L_2 - R'$ is infinite. We claim that there exists a string $w \in L_2$ such that $R_1/\{w\} - R'$ is infinite.

To see this, let $M_1 = (Q_1, \Sigma, \delta_1, q_1, F_1)$ be a DFA for R_1. By the usual construction [24, p. 63] if we let $F_1' = \{q \in Q \ : \ \text{there exists } y \in L_2 \text{ such that } \delta_1(q, y) \in F\}$, then $M_1' = (Q_1, \Sigma, \delta_1, q_1, F_1')$ is a DFA for R_1/L_2. Let $M_2 = (Q_2, \Sigma, \delta_2, q_2, F_2)$ be a DFA for R'. Let $M = (Q_1 \times Q_2, \Sigma, \delta, q_0, F)$ where $q_0 = [q_1, q_2]$, $F = F_1' \times (Q_2 - F_2)$ and $\delta([p,q], a) = [\delta_1(p,a), \delta_2(q,a)]$ for $p \in Q_1, q \in Q_2$, and $a \in \Sigma$. Then M accepts the language $R_1/L_2 - R'$. Since $R_1/L_2 - R'$ is infinite, by the pigeonhole principle, there must be a state $t \in F$ such that $L = \{x \in \Sigma^* \ : \ \delta(q_0, x) = t\}$ is infinite. Suppose $t = [r, s]$. Then $r \in F_1'$, and hence there exists $w \in L_2$ such that $\delta_1(r, w) \in L_2$. It follows that $L \subseteq R_1/\{w\} - R'$, and the result follows.

Now let $B = (R_1/L_2 - R')\{w\} \ \cap \ R_1$. Clearly B is regular and $B \subseteq R_1$. We claim that

$$L_1 \subseteq R_1 - B \subseteq R_1 \tag{3.1.1}$$

and B is infinite.

First, let's prove (3.1.1). Since $L_1 \subseteq R_1$, it suffices to show that $L_1 \ \cap \ B = \emptyset$. Suppose there exists a string x with $x \in L_1$ and $x \in B$. Since $x \in B$, we can write $x = tw$ where $t \in R_1/L_2 - R'$. Hence $t \notin R'$. But $tw \in L_1$, which implies $t \in L_1/\{w\} \subseteq L_1/L_2 \subseteq R'$, so $t \in R'$, a contradiction.

Now let us prove that B is infinite. We know from above that $R_1/\{w\} - R'$ is infinite, say $R_1/\{w\} - R' = \{x_1, x_2, \ldots, \}$. Since $x_i \in R_1/\{w\}$ for each $i \geq 1$, we have $x_i w \in R_1$ for each $i \geq 1$. Also $x_i w \in (R_1/L_2 - R')\{w\}$. Hence $x_i w \in B$. But all the $x_i w$ are distinct and hence B is infinite.

Since $L_1 \subseteq R_1 - B \subseteq R_1$, and $R_1 - B$ is regular, and $R_1 - (R_1 - B) = B$ is infinite, it now follows that R_1 is not a minimal regular cover of L_1, a contradiction.

□

3.1.2 Closure under morphisms of MRC

We now turn to the properties of minimal regular covers under homomorphisms and inverse homomorphisms. We find that forward homomorphisms and ϵ-free inverse homomorphisms both preserve regular covers.

Recall that $\varphi : \Sigma^* \to \Delta^*$ is a homomorphism if $\varphi(xy) = \varphi(x)\varphi(y)$ for all $x, y \in \Sigma^*$. We first have a set-theoretic lemma which will be useful in the proof of Theorem 3.1.7:

Lemma 3.1.6 *Let L_1, L_2 be languages with $L_1 \subseteq \Sigma^*$ and $L_2 \subseteq \Delta^*$. Let $\varphi : \Sigma^* \to \Delta^*$ be a homomorphism. Then*

$$\varphi(L_1) - L_2 \subseteq \varphi(L_1 - \varphi^{-1}(L_2)).$$

Proof: Let $x \in \varphi(L_1) - L_2$. Then $x = \varphi(y)$ for some $y \in L_1$, and $x \notin L_2$. Then $\varphi(y) \notin L_2$. If $y \in \varphi^{-1}(L_2)$, then $\varphi(y) \in \varphi(\varphi^{-1}(L_2)) \subseteq L_2$, a contradiction. Thus $y \notin \varphi^{-1}(L_2)$. But then $y \in L_1 - \varphi^{-1}(L_2)$ and $x = \varphi(y) \in \varphi(L_1 - \varphi^{-1}(L_2))$, as desired. □

Theorem 3.1.7 *Let L be a language with minimal regular cover $R \subseteq \Sigma^*$. Let $\varphi : \Sigma^* \to \Delta^*$ be a homomorphism. Then $\varphi(L)$ has minimal regular cover $\varphi(R)$.*

Proof: Suppose, to arrive at a contradiction, that R_0 is a regular language such that $\varphi(L) \subseteq R_0 \subseteq \varphi(R)$, and $\varphi(R) - R_0$ is infinite. We have $L \subseteq \varphi^{-1}(\varphi(L)) \subseteq \varphi^{-1}(R_0)$.

Define R' to be the language

$$R' = \varphi^{-1}(R_0) \cap R.$$

By the closure properties of regular languages, R' is regular. Also, $L \subseteq \varphi^{-1}(R_0)$ and $L \subseteq R$ imply that $L \subseteq R' \subseteq R$.

By Lemma 3.1.6, we have that

$$\varphi(R) - R_0 \subseteq \varphi(R - \varphi^{-1}(R_0)).$$

Thus, $\varphi(R) - R_0$ infinite implies $\varphi(R - \varphi^{-1}(R_0))$ is also infinite. This implies $R - R' = R - (\varphi^{-1}(R_0) \cap R) = R - \varphi^{-1}(R_0)$ is infinite. This is our contradiction.
□

We prove some preliminary lemmas before establishing the closure of minimal regular covers under inverse ϵ-free homomorphisms. Recall that we define $\varphi^{-1}(x) = \{y \ : \ \varphi(y) = x\}$ for any $x \in \Sigma^*$ and naturally extend this to languages as

$$\varphi^{-1}(L) = \bigcup_{x \in L} \varphi^{-1}(x).$$

We say that a homomorphism $\varphi : \Sigma^* \to \Delta^*$ is ϵ-free if $\varphi(a) \neq \epsilon$ for each $a \in \Sigma$.

Lemma 3.1.8 *Let $\varphi : \Sigma^* \to \Delta^*$ be an ϵ-free homomorphism. Then $S \subseteq \Sigma^*$ infinite implies $\varphi(S)$ infinite.*

Proof: Since S is infinite, there exists a set $\{x_i\}_{i \geq 0} \subseteq S$ such that $|x_i| = n_i$ and $\{n_i\}_{i \geq 0}$ is a monotonic increasing sequence.

Let $m_i = |\varphi(x_i)|$. Since φ is ϵ-free, $m_i \geq n_i$. Then we can choose a monotonic subsequence of $\{m_i\}_{i\geq 0}$, say $\{m_{i_j}\}_{j\geq 0}$. Then $\{\varphi(x_{i_j})\}_{j\geq 0} \subseteq \varphi(S)$ is an infinite set.

□

Given a homomorphism $\varphi : \Sigma^* \to \Delta^*$, define the language $E = E(\varphi)$ by

$$E = \{x \in \Delta^* \ : \ \varphi^{-1}(x) = \emptyset\}.$$

Lemma 3.1.9 *For any $\varphi : \Sigma^* \to \Delta^*$, the language E is regular.*

Proof: Let $\Sigma = \{b_1, b_2, \ldots, b_n\}$. For each i with $1 \leq i \leq n$, let $w_i \in \Delta^*$ be given by $w_i = \varphi(b_i)$. We claim that

$$\overline{E} = (w_1 + w_2 + \cdots + w_n)^*$$

which will establish the lemma.

Let $x \in (w_1 + w_2 + \cdots + w_n)^*$, say $x = w_{i_1} w_{i_2} \cdots w_{i_k}$ with $1 \leq i_j \leq n$ for all $1 \leq j \leq k$. But then $\varphi(b_{i_1} b_{i_2} \cdots b_{i_n}) = x$, thus $\varphi^{-1}(x) \neq \emptyset$ and thus $x \notin E$.

If $x \in \overline{E}$, then there exists $y \in \Sigma^*$ such that $\varphi(y) = x$. Let $y = b_{j_1} b_{j_2} \cdots b_{j_s}$ with $1 \leq j_i \leq n$ for all $1 \leq i \leq s$. But now $x = \varphi(y) = w_{j_1} w_{j_2} \cdots w_{j_s}$. □

Given $\varphi : \Sigma^* \to \Delta^*$ and regular languages $R_1 \subseteq \Delta^*$ and $R_2 \subseteq \Sigma^*$, define $T = T(\varphi, R_1, R_2)$ as

$$T = \{x \in \varphi^{-1}(R_1) - R_2 \ : \ \varphi(x) \in \varphi(R_2)\}.$$

We establish the following facts about T:

Lemma 3.1.10 *Let $\varphi : \Sigma^* \to \Delta^*$ be a homomorphism and let $R_1 \subseteq \Sigma^*$ and $R_2 \subseteq \Sigma^*$ be regular languages. Let $T = T(\varphi, R_1, R_2)$. Then*

(a). T is regular;

(b). For any language $L \subseteq R_1$ such that $\varphi^{-1}(L) \subseteq R_2$, $\varphi(T) \cap L = \emptyset$.

Proof: (a). Let $T' = \{x \in \Sigma^* : \varphi(x) \in \varphi(R_2)\}$. Note that $T' = \varphi^{-1}(\varphi(R_2))$. Thus T' is regular. Also,

$$T = T' \cap (\varphi^{-1}(R_1) - R_2),$$

giving that T is regular.

(b). Let L be a language with $L \subseteq R_1$ and $\varphi^{-1}(L) \subseteq R_2$. Let $w \in L \cap \varphi(T)$. By definition of T, $T \subseteq \varphi^{-1}(R_1) - R_2$. Thus, $w \in \varphi(T)$ implies that there exists $u \in \varphi^{-1}(R_1) - R_2$ such that $\varphi(u) = w$. Then using the fact that $w = \varphi(u) \in L$, $u \in \varphi^{-1}(L) \subseteq R_2$. This contradicts our choice of $u \in \varphi^{-1}(R_1) - R_2$. $\square$

We now prove that minimal regular covers are closed under ϵ-free inverse homomorphism.

Theorem 3.1.11 *Let $\varphi : \Sigma^* \to \Delta^*$ be an ϵ-free homomorphism. Let $L, R \subseteq \Delta^*$ be languages with R a minimal regular cover for L. Then $\varphi^{-1}(L)$ has minimal regular cover $\varphi^{-1}(R)$.*

Proof: The inclusion $\varphi^{-1}(L) \subseteq \varphi^{-1}(R)$ is trivial. Let R' be a regular language such that $\varphi^{-1}(L) \subseteq R' \subseteq \varphi^{-1}(R)$. We must show that $\varphi^{-1}(R) - R'$ is finite.

Let $E \subseteq \Sigma^*$ be $E = E(\varphi)$ as defined above. Let $R_0 \subseteq \Sigma^*$ be the regular language

$$R_0 = \varphi(R') \cup (R \cap E).$$

R_0 is regular by Lemma 3.1.9 and our standard closure properties of regular languages.

We claim that

$$L \subseteq R_0 \subseteq R. \tag{3.1.2}$$

The inclusion $R_0 \subseteq R$ is true by definition of R_0 and the fact that $\varphi(R') \subseteq R$. Now, let $x \in L - \varphi(R')$. If there is $y \in \Sigma^*$ with $\varphi(y) = x$, then $y \in \varphi^{-1}(x) \subseteq \varphi^{-1}(L) \subseteq R'$. So $x \in \varphi(R')$, contrary to our choice of x. Thus $\varphi^{-1}(x) = \emptyset$ and so $x \in E$. Since $x \in L \subseteq R$, we get that $x \in R \cap E$. Thus, $L \subseteq R_0 = \varphi(R') \cup (R \cap E)$.

Let $T = T(\varphi, R, R')$ be as defined above. Then by Lemma 3.1.10, T is regular and $\varphi(T) \cap L = \emptyset$. From (3.1.2), we have seen above that $L \subseteq R_0$. Thus, $L \subseteq (R_0 - \varphi(T))$. Also from (3.1.2), $R_0 \subseteq R$. Thus

$$L \subseteq R_0 - \varphi(T) \subseteq R.$$

But since R is a minimal regular cover of L, we must have that

$$R - (R_0 - \varphi(T)) \text{ is finite.} \tag{3.1.3}$$

Lemma 3.1.12 *Given the conditions as above,*

$$\varphi(\varphi^{-1}(R) - R') \subseteq R - (R_0 - \varphi(T)).$$

Proof: Let $x \in \varphi(\varphi^{-1}(R) - R')$. Then, there exists a $y \in \varphi^{-1}(R) - R'$ such that $\varphi(y) = x$. Now $y \in \varphi^{-1}(R)$ implies that $x = \varphi(y) \in R$. In order to obtain a contradiction, assume $x \in R_0 - \varphi(T)$. Since $x \in R_0$, either $x \in E$ or $x \in \varphi(R')$, by definition of R_0.

If $x \in E$ then $\varphi^{-1}(x) = \emptyset$. However, we know that $y \in \varphi^{-1}(x)$. Therefore, we must have $x \in \varphi(R')$. That is, $x = \varphi(y) \in \varphi(R')$. But recalling our definition of T, we note that $y \in \varphi^{-1}(R) - R'$ and $\varphi(y) \in \varphi(R')$ implies $y \in T$ and $x = \varphi(y) \in \varphi(T)$. But this is contrary to our assumption on x. Thus $x \in R - (R_0 - \varphi(T))$. □

But now by Lemma 3.1.12 and (3.1.3), we have that $\varphi(\varphi^{-1}(R) - R')$ is finite. So, since φ is ϵ-free, we must have that $\varphi^{-1}(R) - R'$ is also finite, by Lemma 3.1.8. This proves the result. □

Thus, we may now easily construct languages which are unbounded but which have a minimal regular cover. This complements Section 3.3, which will consider exclusively bounded languages with minimal regular covers.

Example 3.1.13: Let Σ be an arbitrary alphabet and let $\varphi : \Sigma^* \to 0^*$ be the homomorphism defined by $\varphi(a) = 0$ for all $a \in \Sigma$. Let L be any language with minimal regular cover 0^*. By Theorem 3.3.5 $I(L) = \{n \ : \ 0^n \in L\}$ intersects every arithmetic progression.

Then $\varphi^{-1}(L) = \{x \in \Sigma^* \ : \ |x| \in I(L)\}$ has minimal regular cover Σ^*. If $|\Sigma| > 1$, then $\varphi^{-1}(L)$ is unbounded. Thus, for any set I which intersects every arithmetic progression, the language $L_I = \{x \in \Sigma^* \ : \ |x| \in I\}$ has minimal regular cover Σ^*.

For example, $L = \{x \in \{a,b\} \ : \ |x| \neq n + n!, \ \forall n \geq 1\}$ has minimal regular cover $\{a,b\}^*$. □

Following Theorem 3.1.11, we consider the closure properties of languages with minimal regular covers under inverse ϵ-free substitution. A substitution $\varphi : \Sigma^* \rightarrow 2^{\Delta^*}$ is ϵ-free if $\epsilon \notin \varphi(a)$ for all $a \in \Sigma$ (see, eg. [44, p. 19]). Recall that for a language $L \subseteq \Delta^*$ and a substitution $\varphi : \Sigma^* \rightarrow 2^{\Delta^*}$, we may define the inverse image of a language L by

$$\varphi^{-1}(L) = \{x \in \Sigma^* \ : \ \varphi(x) \subseteq L\}. \tag{3.1.4}$$

Then Theorem 3.1.11 is not true if we replace homomorphism by substitution, even regular substitution. To see this, let $L = \{0^n \ : \ n \geq 1 \text{ not prime}\}$. Anticipating Lemma 3.2.2, L has minimal regular cover 0^+. Define the substitution $\varphi : \{a,b\}^* \rightarrow 2^{0^*}$ by $\varphi(a) = 0^+$ and $\varphi(b) = 00(00)^*$. Note that $\varphi(b^i) = (00)^i(00)^* \subseteq L$ and that if $|x|_a \neq 0$, $\varphi(x) = 0^+ \nsubseteq L$. Then $\varphi^{-1}(L) = b^+$ while $\varphi^{-1}(0^+) = (a+b)^+$. So $\varphi^{-1}(L)$ does not have minimal regular cover $\varphi^{-1}(0^+)$.

However, we may also use the definition of inverse substitution given by

$$\varphi^{[-1]}(L) = \{x \in \Sigma^* \ : \ \varphi(x) \cap L \neq \emptyset\}.$$

We use the notation $\varphi^{[-1]}$ to distinguish it from the definition of inverse substitution given by (3.1.4). With this in mind, we prove that if L is a language with minimal regular cover 0^*, inverse ϵ-free substitution preserves minimal regular covers. We start with some preliminary lemmas, including a proof that minimal regular covers are closed under finite inverse ϵ-free substitution.

Lemma 3.1.14 *Let $L \subseteq 0^*$ have minimal regular cover 0^*. Let Σ be an alphabet and $\varphi : \Sigma^* \to 2^{0^*}$ be a regular ϵ-free substitution. For all $x \in \Sigma^*$ such that $|\varphi(x)|$ is infinite, $x \in \varphi^{[-1]}(L)$.*

Proof: Let $x \in \Sigma^*$ such that $\varphi(x)$ is infinite. Since φ is a regular substitution, $\varphi(x)$ is regular. Thus we know by Lemma 3.2.1 that $\varphi(x) \supseteq 0^r(0^s)^*$ for some integers $r \geq 0$ and $s \geq 1$. Since φ is ϵ-free, $s \neq 0$. We consider the arithmetic progression $r + sm$ for $m \geq 0$. Anticipating Theorem 3.3.5, there exists an m such that $0^{r+sm} \in L$. Thus, $\varphi(x) \cap L \neq \emptyset$. Thus $x \in \varphi^{[-1]}(L)$. □

Theorem 3.1.15 *Let $L \subseteq \Delta^*$ be a language with minimal regular cover R. Let $\varphi : \Sigma^* \to 2^{\Delta^*}$ be a finite ϵ-free substitution. Then $\varphi^{[-1]}(L)$ has minimal regular cover $\varphi^{[-1]}(R)$.*

Proof: Let $\Sigma = \{a_1, \ldots, a_n\}$. For each $1 \leq i \leq n$, let m_i be integers such that $|\varphi(a_i)| = m_i$. Let $\varphi(a_i) = \{u_{i,1}, \ldots, u_{i,m_i}\}$ for words $u_{i,j} \in \Delta^*$. Now, let $m = \prod_{i=1}^{n} m_i$. Now, we define m homomorphisms $\varphi_j : \Sigma^* \to \Delta^*$ for $1 \leq j \leq m$, such that for each $(u_{1,j_1}, u_{2,j_2}, \ldots, u_{n,j_n}) \in \varphi(a_1) \times \varphi(a_2) \times \ldots \times \varphi(a_n)$, there exists a j with $1 \leq j \leq m$ such that $\varphi_j(a_i) = u_{i,j_i}$. Note that as φ is ϵ-free, each φ_j is also ϵ-free. We claim that

$$\bigcup_{j=1}^{m} \varphi_j^{-1}(L) = \varphi^{[-1]}(L) \tag{3.1.5}$$

We first prove the left-to-right inclusion of (3.1.5): Let $x \in \varphi_j^{-1}(L)$. Then $\varphi_j(x) \in L$, since φ_j is a homomorphism. But now $\varphi_j(x) \subseteq \varphi(x)$, by definition of φ_j, so that $\varphi(x) \cap L \neq \emptyset$ and $x \in \varphi^{[-1]}(L)$.

To prove the right-to-left inclusion of (3.1.5), let $x \in \varphi^{[-1]}(L)$. Then $\varphi(x) \cap L \neq \emptyset$, say in particular that $z \in \varphi(x) \cap L$. But now if $x = a_{k_1} \cdots a_{k_r}$ for $a_{k_i} \in \Sigma$, then there is a decomposition of z as $z_{k_1} \cdots z_{k_r}$ such that $z_{k_i} \in \varphi(a_{k_i})$. Now, by the definition of φ_j, there exists a j such that $\varphi_j(x) = z$. Then $z \in L$ implies $x \in \varphi_j^{-1}(L)$. Thus, the inclusion is proved.

Now, by Theorems 3.1.1 and 3.1.11, $\bigcup_{j=1}^m \varphi_j^{-1}(L)$ has minimal regular cover $\bigcup_{j=1}^m \varphi_j^{-1}(R)$. By the equality of (3.1.5), applied with R in place of L, the theorem is proved. □

Theorem 3.1.16 *Let $L \subseteq 0^*$ be a language with minimal regular cover 0^*. Let $\varphi : \Sigma^* \to 2^{0^*}$ be a regular ϵ-free substitution. Then $\varphi^{[-1]}(L)$ has minimal regular cover Σ^*.*

Proof: Let $\Sigma = \{a_1, a_2, \ldots, a_n\}$ and let $\varphi(a_i) = R_i$ for all $1 \leq i \leq n$. Define subsets X, Y of Σ as

$$X = \{a_i \ : \ R_i \text{ is infinite }\},$$

$$Y = \{a_i \ : \ R_i \text{ is finite }\}.$$

Clearly $X \cap Y = \emptyset$ and $X \cup Y = \Sigma$. We claim that the following equalities hold:

$$\varphi^{[-1]}(L) \cap \overline{Y^*} = \overline{Y^*} \qquad (3.1.6)$$

$$\varphi^{[-1]}(L) \cap Y^* = \varphi|_Y^{[-1]}(L). \tag{3.1.7}$$

Consider first (3.1.6). It is equivalent to $\overline{Y^*} \subseteq \varphi^{[-1]}(L)$. Thus, consider $x \in \overline{Y^*}$, which implies that $|x|_X > 0$. So, $\varphi(x)$ is infinite, and by Lemma 3.1.14, we have that $\varphi(x) \cap L \neq \emptyset$ and $x \in \varphi^{[-1]}(L)$. Thus (3.1.6) is established.

As to (3.1.7), let $x \in \varphi|_Y^{[-1]}(L)$. Then $\varphi|_Y(x) \cap L \neq \emptyset$. But $\varphi|_Y$ is only defined on Y^*, so $x \in Y^*$. Also $\varphi|_Y(x) \cap L \subseteq \varphi(x) \cap L$. So $x \in \varphi^{[-1]}(L)$. Conversely, let $x \in \varphi^{[-1]}(L) \cap Y^*$. Then $x \in Y^*$, so $\varphi|_Y(x)$ is defined and in fact is equal to $\varphi(x)$. Thus clearly $x \in \varphi|_Y^{[-1]}(L)$. This establishes the inclusion.

Now $\varphi|_Y : Y \to 2^{0^*}$ is a finite substitution, so by Theorem 3.1.15, $\varphi|_Y^{[-1]}(L)$ has minimal regular cover $\varphi|_Y^{[-1]}(0^*)$. Further,

$$\begin{aligned} \varphi^{[-1]}(L) &= (\varphi^{[-1]}(L) \cap \overline{Y^*}) \cup (\varphi^{[-1]}(L) \cap Y^*) \\ &= \overline{Y^*} \cup \varphi|_Y^{[-1]}(L) \end{aligned}$$

has minimal regular cover $\overline{Y^*} \cup \varphi|_Y^{[-1]}(0^*)$.

But now (3.1.6) and (3.1.7) give us that $\varphi^{[-1]}(0^*) = \overline{Y^*} \cup \varphi|_Y^{[-1]}(0^*)$. Thus $\varphi^{[-1]}(L)$ has minimal regular cover $\varphi^{[-1]}(0^*)$. The result follows on noting that necessarily $\varphi^{[-1]}(0^*) = \Sigma^*$. □

One particular case is worth noting as a corollary:

Corollary 3.1.17 *Let $L \subseteq 0^*$ be a language with minimal regular cover 0^*. Let $\varphi : \Sigma^* \to 2^{0^*}$ be a regular ϵ-free substitution such that for all $a \in \Sigma$, $\varphi(a)$ is infinite. Then $\varphi^{[-1]}(L) = \Sigma^*$.*

Proof: The corollary follows on letting $Y = \emptyset$ in the proof of Theorem 3.1.16. $\square$

We have a partial generalization of Theorem 3.1.16. It follows from the following lemma:

Lemma 3.1.18 *Let L have minimal regular cover R. Then $L \cup \overline{R}$ has minimal regular cover Σ^*.*

Proof: Let $L \cup \overline{R} \subseteq R' \subseteq \Sigma^*$. We must show $\overline{R'}$ is finite. Now $\overline{R'} \subseteq \overline{L} \cap R \subseteq R$.

Also, $(L \cup \overline{R'}) \cap R \subseteq R' \cap R$, and this cover is minimal, by Theorem 3.1.4. So by Theorem 3.0.1, $R - (R' \cap R)$ is finite. That is $R \cap \overline{R'}$ is finite. But since $\overline{R'} \subseteq R$, $R \cap \overline{R'} = \overline{R'}$. So $\overline{R'}$ is finite, as required. $\square$

The following is an easy lemma, stated without proof:

Lemma 3.1.19 *Let φ be a substitution. Then $\varphi^{[-1]}(A \cup B) = \varphi^{[-1]}(A) \cup \varphi^{[-1]}(B)$ and $A \subseteq B$ implies $\varphi^{[-1]}(A) \subseteq \varphi^{[-1]}(B)$.*

Theorem 3.1.20 *Let $L \subseteq 0^*$ be a language with minimal regular cover R. Let $\varphi : \Sigma^* \to 2^{0^*}$ be a regular ϵ-free substitution. Then $\varphi^{[-1]}(L) - \varphi^{[-1]}(\overline{R})$ has minimal regular cover $\varphi^{[-1]}(R) - \varphi^{[-1]}(\overline{R})$.*

Proof: By Lemma 3.1.18, we know that $L \cup \overline{R}$ has minimal regular cover 0^*. Thus Theorem 3.1.16 implies that $\varphi^{[-1]}(L \cup \overline{R})$ has minimal regular cover Σ^*.

By Lemma 3.1.19, $\varphi^{[-1]}(L) \cup \varphi^{[-1]}(\overline{R})$ has minimal regular cover Σ^* and thus by Theorem 3.1.4 and Lemma 3.1.19 again, $\varphi^{[-1]}(L) \cup (\varphi^{[-1]}(\overline{R}) \cap \varphi^{[-1]}(R))$ has minimal regular cover $\varphi^{[-1]}(R)$. Since $\varphi^{[-1]}(\overline{R}) \cap \varphi^{[-1]}(R)$ is regular and $L \subseteq R$, we can apply again Theorem 3.1.4 to yield our final result: $\varphi^{[-1]}(L) - \varphi^{[-1]}(\overline{R})$ has minimal regular cover $\varphi^{[-1]}(R) - \varphi^{[-1]}(\overline{R})$. □

3.2 Non-closure properties of MRCs

We examine operations which are not closed under minimal regular cover. We begin with a lemma which will be useful below. It is due to Shallit:

Lemma 3.2.1 *If $L \subseteq 0^*$ is a regular language, then there exist integers $m \geq 0$, $n \geq 1$ and sets $A \subseteq \{\epsilon, 0, \ldots, 0^{m-1}\}$ and $B \subseteq \{0^m, 0^{m+1}, \ldots, 0^{m+n-1}\}$ such that $L = A \cup B(0^n)^*$. Furthermore, L is infinite if and only if $B \neq 0$.*

Proof: If L is regular, then it is accepted by a DFA. The structure of this DFA must consists of a "tail" of $m \geq 0$ states followed by a cycle of $n \geq 1$ states. The result easily follows. □

We have the following useful lemma, due to Shallit:

Lemma 3.2.2 *Suppose $a_1 < a_2 < a_3 < \cdots$ is a strictly increasing sequence of non-negative integers such that $a_n/n \to \infty$ as $n \to \infty$. Let $A = \{0^{a_i} \; : \; i \geq 0\}$. Then $L = 0^* - A$ is minimally covered by 0^*.*

Proof: We have $0^* - L = A$. If A contained an infinite regular subset, then by Lemma 3.2.1, there would exist integers $i \geq 0$ and $j \geq 1$ such that $0^i(0^j)^* \subseteq A$. But then there would be an $\epsilon > 0$ such that $a_n/n \leq j + \epsilon$ for all n sufficiently large, a contradiction. □

We now show that there exists a language L such that L has a minimal regular cover, while L^* has no minimal regular cover. Since if L is unary then L^* is regular (see [24, Exercise 3.7, p. 72]), we must consider a alphabet of size strictly greater than one:

Theorem 3.2.3 *Let* $D = \{0^{n+n!} \ : \ n \geq 0\}$. *Then* $L = 1\overline{D}$ *has minimal regular cover* 10^*, *but* L^* *has no minimal regular cover.*

Proof: By Lemma 3.2.2, $\overline{D}$ has minimal regular cover 0^*. By Lemma 3.1.3, L has minimal regular cover 10^*.

Now, we establish that L^* has no minimal regular cover. Let R be a regular language with $L^* \subseteq R$. Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA with r states such that $L(M) = R$. Let n be an integer with $n > r$ and n not of the form $m + m!$ for all $m \geq 1$.

Consider 10^n. Since n is not of the form $m + m!$, $10^n \in L \subseteq L^*$. Further, since $L^* \subseteq R$, 10^n is accepted by M and thus $\delta(q_0, 10^n) = q_1$ for some state $q_1 \in F$. But by the pigeonhole principle, since $n > r$ there exists some state q_2 such that q_2 is visited twice in the computation of 10^n. Let the distance between successive visits to q_2 be k, where k is an integer with $1 \leq k \leq r$. Thus, since we may now repeat the computation between successive visits of q_2 any number of times, it follows that

$$\delta(q_0, 10^{n+jk}) = q_1$$

for all $j \geq 0$. Since $k \leq r < n$, $j = n!/k$ is an integer, and so

$$\delta(q_0, 10^{n+n!}) = q_1. \tag{3.2.1}$$

Now $1000 \in L$. Thus $10^n(1000)^i \in L^*$ for all $i \geq 0$. Since $\delta(q_0, 10^n) = q_1$ and $L^* \subseteq R$, $\delta(q_1, (1000)^i) \in F$ for all $i \geq 0$. Combining this with (3.2.1) gives $\delta(q_0, 10^{n+n!}(1000)^i) \in F$ for all $i \geq 0$ and so $10^{n+n!}(1000)^* \subseteq R$ for our choice of n. But $10^{n+n!}(1000)^i \notin L$ for all $i \geq 0$, thus

$$L^* \subseteq R - \{10^{n+n!}(1000)^i \ : \ i \geq 0\} \subseteq R.$$

Thus R is not a minimal regular cover of L^*. Thus L^* has no minimal regular cover.

□

We now turn to arbitrary intersection. We see that two languages may possess minimal regular covers, while their intersection may not. The theorem is proved after establishing that the unary language $\{0^{n^2} \ : \ n \geq 1\}$ has no minimal regular cover. We introduce the following notation: $\omega(n)$ denotes the total number of prime factors of n, counted with multiplicity. For example, $\omega(2^3 5^7 11^8) = 18$. The entire argument is due to Shallit:

Lemma 3.2.4 *Let $L_s = \{0^{n^2} \ : \ n \geq 1\}$. Then L_s has no minimal regular cover.*

Proof: Suppose R is a minimal regular cover of L_s. Then by Lemma 3.2.1 we can write $R = A \cup B(0^n)^*$ with $B \subseteq \{0^m, 0^{m+1}, \ldots, 0^{m+n-1}\}$. Let p be an odd prime not dividing n, and let a be a quadratic non-residue (mod p), i.e., the Legendre symbol $\left(\frac{a}{p}\right) = -1$. Since L_s is infinite, R must be infinite, and so $B \neq \emptyset$. Choose a b such that $0^b \in B$. Let r be an integer, $m \leq r < m + pn$ such that $r \equiv b(\text{mod } n)$ and $r \equiv a(\text{mod } p)$; such an r exists by the Chinese remainder theorem.

Now let $T = (0^{pn})^* 0^r$. Note that $T \subseteq R$. If $0^c \in T$, then c is not a square, since $\left(\frac{c}{p}\right) = \left(\frac{r}{p}\right) = \left(\frac{a}{p}\right) = -1$. It follows that $L_s \subseteq R - T$, so $R - T$ is a regular cover of L_s. On the other hand, $R - (R - T) = T$ is infinite, so R is not minimal, a contradiction. $\square$

Theorem 3.2.5 *Let*

$$L_0 = \{0^n \ : \ \omega(n) \equiv 0(mod\ 2)\}$$

and

$$L_1 = \{0^n \ : \ n \text{ is a square or } \omega(n) \equiv 1(mod\ 2)\}.$$

Then Σ^* is a minimal regular cover of both L_0 and L_1, but $L_0 \cap L_1$ has no minimal regular cover.

Proof: First, consider L_0. By Theorem 3.3.5, it suffices to show that $I(L_0) = \{n : \omega(n) \equiv 2(\text{mod } 2)\}$ intersects every arithmetic progression. Let $A_{n,a} := \{nt + a : t \geq 0\}$ be an arithmetic progression. Let $g = \gcd(a, n)$; note that if $a = 0$ then $g = n$. Then by Dirichlet's theorem there exists infinitely many prime numbers $p \equiv a/g(\text{mod } n/g)$, so in particular there exists such a p with $p \geq n$.

Now there are two cases to consider. If $\omega(g) \equiv 1(\text{mod } 2)$, then $\omega(gp) \equiv 0(\text{mod } 2)$. Also $gp \equiv a(\text{mod } n)$ and $gp \geq n$, so $gp \in A_{n,a}$ and hence $I(L_0)$ intersects $A_{n,a}$.

If $\omega(g) \equiv 0(\text{mod } 2)$, then by Dirichlet's theorem there exists a prime $q \equiv 1(\text{mod } n/g)$. Then $\omega(gpq) \equiv 0(\text{mod } 2)$. Also $gpq \equiv a(\text{mod } n)$ and $gpq \geq n$, so $gpq \in A_{n,a}$ and hence $I(L_0)$ intersects $A_{n,a}$. It follows that 0^* is a minimal regular cover of L_0.

A similar argument applies to $\{0^r \ : \ \omega(r) \equiv 1 (\text{mod } 2)\}$, and hence, by Theorem 3.1.2, L_1 has 0^* as a minimal regular cover. However, $L_0 \cap L_1 = \{0^{n^2} \ : \ n \geq 1\} = L_s$, and L_s has no minimal regular cover by Lemma 3.2.4. $\square$

Lemma 3.1.3 demonstrated that minimal regular covers were closed under concatenation with finite languages. However, the following theorem, due to the author and Shallit, shows that this closure fails when we consider even regular languages.

Theorem 3.2.6 *Let $L_1 = 1^+ - \{1^p \ : \ p \text{ prime}\}$. and $L_2 = 0^*$. Then L_1 has minimal regular cover 1^+ and L_2 has minimal regular cover itself. However, L_1L_2 has no minimal regular cover.*

Proof: Let $L = L_1L_2$, and suppose R is a minimal regular cover for L. Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA with $L(M) = R$. Let $n = |Q|$, and p be a prime with $p > n$. Since $1^{p^2}0^* \subseteq L$, we have $\delta(q_0, 1^{p^2}0^j) \in F$ for all $j \geq 0$. However, by the pigeonhole principle, there must be some state such that $\delta(q_0, 1^a) = \delta(q_0, 1^{a+b})$ for some $a \geq 0$,$b \geq 1$ such that $a + b \leq n$. We can go around this loop as many times as we want, so it follows that $\delta(q_0, 1^{p^2+(i-1)b}0^j) \in F$ for all $i, j \geq 0$. By Dirichlet's theorem, there exists i such that $p^2 + (i-1)b$ is a prime number. Call this prime number q. Let $R' = R - 1^q0^*$. Then $L \subseteq R' \subseteq R$ and $R - R'$ is infinite. This shows that L has no minimal regular cover. $\square$

We now consider closure under inverse homomorphism. The next lemma shows that minimal regular covers are not closed under arbitrary inverse homomorphism. We have already seen that this closure holds if we restrict ourselves to ϵ-free inverse homomorphism.

Theorem 3.2.7 *Let* $L = \{0^{n+n!} \ : \ n \geq 1\}$ *and* $\varphi : \{a, b\} \to \{0\}$ *be the homomorphism defined by* $\varphi(a) = 0$ *and* $\varphi(b) = \epsilon$*. Then* $\overline{L}$ *has minimal regular cover* 0^**, but* $\varphi^{-1}(\overline{L})$ *has no minimal regular cover.*

Proof: By Lemma 3.2.2, $\overline{L}$ has minimal regular cover 0^*. Define $L' \subseteq \{a, b\}^*$ as $L' = \varphi^{-1}(\overline{L})$. Observe that

$$L' = \{x \in \{a, b\}^* \ : \ |x|_a \neq m + m! \ \forall m \geq 1\}.$$

We claim that L' has no minimal regular cover. Let R be a minimal regular cover, so that $L' \subseteq R \subseteq \{a, b\}^*$. Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA with r states such that $L(M) = R$. Let n be an integer with $n > r$ and n not of the form $m + m!$ for all $m \geq 1$.

Since $L' \subseteq R$, and $a^n \in L'$, we have $\delta(q_0, a^n) = q_1$ for some $q_1 \in F$. But $a^n b^i \in L'$ for all $i \geq 0$, so we must have $\delta(q_1, b^i) \in F$ for all $i \geq 0$. But the pigeonhole principle implies that there is an integer k, with $1 \leq k \leq r$ such that $\delta(q_0, a^{n+jk}) = \delta(q_0, a^n) = q_1$ for all $j \geq 0$. Choose $j = n!/k$ to yield $\delta(q_0, a^{n+n!}) = q_1$. Thus for all $i \geq 0$, $\delta(q_0, a^{n+n!}b^i) \in F$ and so $a^{n+n!}b^* \subseteq R$ for our choice of n. But now $a^{n+n!}b^i \notin L'$ for all $i \geq 0$ and so $L' \subseteq R - a^{n+n!}b^* \subseteq R$. So R is not a minimal regular cover of L'. □

As a consequence of the proof of Theorem 3.2.6, we can show that minimal regular covers are not closed under substitution, even regular substitution. Recall that $\varphi : \Sigma \to 2^{\Delta^*}$ is a substitution if $\varphi(xy) = \varphi(x)\varphi(y)$ for all $x, y \in \Sigma^*$. A substitution is regular if $\varphi(a)$ is a regular language for all $a \in \Sigma$.

Theorem 3.2.8 *Let* $\varphi : \{0,1\}^* \to 2^{\{0,1\}^*}$ *be the regular substitution defined by* $\varphi(0) = 0^*$ *and* $\varphi(1) = 1$. *Let* $L_1 = 1^+ - \{1^p : p \text{ is prime }\}$. *Then* $L_1 0$ *has minimal regular cover* 1^+0 *but* $\varphi(L_1 0)$ *has no minimal regular cover.*

Proof: We know that $L_1 0$ has minimal regular cover 1^+0 by Lemma 3.2.2 and Theorem 3.1.3. But now $\varphi(L_1 0) = L_1 0^*$, and by Theorem 3.2.6, $L_1 0^*$ has no minimal regular cover. □

We define (arbitrary) shuffle as follows:

$$x \text{ Ш } y = \{x_1 y_1 x_2 y_2 \cdots x_n y_n \ : \ x = x_1 x_2 \cdots x_n,\ y = y_1 y_2 \cdots y_n,\ x_i, y_i \in \Sigma^*\}.$$

Note that we do not insist x_i, y_i be letters from Σ. We extend this to languages as

$$L_1 \text{ Ш } L_2 = \bigcup_{x \in L_1, y \in L_2} x \text{ Ш } y.$$

Similarly, we define the perfect shuffle operation for two strings $x, y \in \Sigma^*$ with $x = x_1 x_2 \cdots x_n$, $y = y_1 y_2 \cdots y_n$, $x_i, y_j \in \Sigma$, as $x \text{ Ш } y = x_1 y_1 x_2 y_2 \cdots x_n y_n$. Note that we insist that x_i, y_i be letters in Σ and $x \text{ Ш } y$ is not defined if $|x| \neq |y|$. We similarly extend the perfect shuffle operation to languages L_1 and L_2 as

$$L_1 \text{ Ш } L_2 = \{x \text{ Ш } y \ : \ x \in L_1, y \in L_2\}.$$

It is relatively easy, as we shall see, to show that that minimal regular covers are not closed under arbitrary shuffle. But we will also see that the same is true of perfect shuffle.

Theorem 3.2.9 *Let* $D = \{0^{n+n!} \ : \ n \geq 0\}$ *and* $L_1 = 1^*$. *Then* $L_0 = \overline{D}$ *has minimal regular cover* 0^*, *and* L_1 *has minimal regular cover itself. But* $L_0 ⧢ L_1$ *has no minimal cover.*

Proof: That L_0 has minimal regular cover 0^* follows from Lemma 3.2.2. It is evident that

$$L_0 ⧢ L_1 = \{x \in \{0,1\}^* \ : \ |x|_0 \neq n + n! \ \forall n \geq 0\}$$

and so $L_0 ⧢ L_1 = L'$ where L' is from the proof of Theorem 3.2.7. Thus, $L_0 ⧢ L_1$ has no minimal regular cover. □

Theorem 3.2.10 *Let*

$$\begin{aligned} L_0 &= \{0^n \ : \ \omega(n) \equiv 0 (mod\ 2)\} \\ L_1 &= \{1^n \ : \ n \text{ is a square or } \omega(n) \equiv 1 (mod\ 2)\}. \end{aligned}$$

Then 0^* *is a minimal regular cover of* L_0 *and* 1^* *is a minimal regular cover of* L_1, *but* $L_0 ⧢ L_1$ *has no minimal regular cover.*

Proof: We know from Theorem 3.2.5 that L_0 and L_1 have minimal regular cover 0^* and 1^*, respectively.

Assume that $L_0 ⧢ L_1$ had minimal regular cover R. Note that $L_0 ⧢ L_1 = \{(01)^r \ : \ r \text{ is a square }\}$. Then letting $h(1) = \epsilon$ and $h(0) = a$, Theorem 3.1.7 gives $h(L_0)$ has minimal regular cover $h(R)$. Since $h(L_0) = L_s$ from Lemma 3.2.4, we have a contradiction. Thus, $L_0 ⧢ L_1$ has no minimal regular cover. □

3.3 Bounded Languages and MRCs

Recall that a language L is **bounded** if there exist words $w_1, w_2, \ldots, w_n$ such that $L \subseteq w_1^* w_2^* \cdots w_n^*$. In this section, we characterize completely bounded languages which possess a minimal regular cover. We begin by giving Theorem 3.3.2, a characterization of bounded regular languages originally due to Ginsburg and Spanier [18]. We first require some preliminary lemmas:

Lemma 3.3.1 *Let $L \subseteq w_1^* w_2^* \cdots w_n^*$ be a bounded regular language. Let L_1 be an arbitrary language. Then*

$$L' = \{w_n^{j_n} \; : \; \exists w_1^{j_1} \cdots w_{n-1}^{j_{n-1}} \in L_1 \textit{ such that } w_1^{j_1} w_2^{j_2} \cdots w_{n-1}^{j_{n-1}} w_n^{j_n} \in L\}$$

is a bounded regular language.

Proof: Clearly $L' \subseteq w_n^*$, thus L' is bounded.

Let $M_1 = (Q_1, \Sigma, \delta_1, q_{0,1}, F_1)$ and $M_2 = (Q_2, \Sigma, \delta_2, q_{0,2}, F_2)$ be DFAs such that $L(M_1) = w_n^*$ and $L(M_2) = L$. Define the NFA $M_2' = (Q_2', \Sigma, \delta_2', q_{0,1}', F_2)$ where $Q_2' = Q_2 \cup q_{0,2}'$ and $\delta_2'(q, a) = \delta_2(q, a)$ for all $q \in Q_2$ and $a \in \Sigma$ and

$$\delta_2'(q_{0,2}', \epsilon) = \{q \in Q_2 \; : \; \exists w_1^{j_1} \cdots w_{n-1}^{j_{n-1}} \in L_1 \text{ such that } \delta_2(q_{0,2}, w_1^{j_1} \cdots w_{n-1}^{j_{n-1}}) = q\}.$$

We now define the NFA $M = (Q, \Sigma, \delta, q_0, F)$ as follows:

$$Q = Q_1 \times Q_2'$$

$$\delta([q', q''], a) = [\delta_1(q', a), \delta_2'(q'', a)]$$

$$q_0 = [q_{0,1}, q'_{0,2}]$$

and define F by

$$F = F_1 \times F_2.$$

We claim that $L(M) = L'$. Let $x \in L(M)$. Then $\delta'_2(q'_{0,2}, x) \in F$. But $\delta(q'_{0,2}, x) = \{\delta(q, x) \ : \ \exists w_1^{j_1} \cdots w_{n-1}^{j_{n-1}} \in L_1$ such that $\delta_2(q_{0,2}, w_1^{j_1} \cdots w_{n-1}^{j_{n-1}}) = q\}$. Also, $x \in w_n^*$, say $x = w_n^{j_n}$. Then, we must have $\delta_2(q_{0,2}, w_1^{j_1} \cdots w_{n-1}^{j_{n-1}} w_n^{j_n}) \in F$, that is, $x \in L'$.

Conversely, let $w_n^{j_n} \in L'$. Then there exists $w_1^{j_1} \cdots w_{n-1}^{j_{n-1}} \in L_1$ such that $w_1^{j_1} \cdots w_{n-1}^{j_{n-1}} w_n^{j_n} \in L$ Let $q \in Q_2$ be the state with $\delta_2(w_1^{j_1} \cdots w_{n-1}^{j_{n-1}}) = q$. Then since $w_1^{j_1} \cdots w_{n-1}^{j_{n-1}} w_n^{j_n} \in L$, $\delta_2(q, w_n^{j_n}) \in F_2$. Thus, by definition of M, $w_n^{j_n} \in L(M)$. Thus $L(M) = L'$ and L' is regular, as required. □

Theorem 3.3.2 *Let $L \subseteq w_1^* w_2^* \cdots w_n^*$ be a bounded language. Then L is regular if and only if there exist an integer N and indices $a_{i,j}, b_{i,j}$ for $1 \le i \le N, 1 \le j \le n$ such that*

$$L = \bigcup_{i=1}^{N} w_1^{a_{i,1}} (w_1^{b_{i,1}})^* \cdots w_n^{a_{i,n}} (w_n^{b_{i,n}})^*.$$

Before proving our theorem, we note the theorem of Ginsburg and Spanier [18]:

Theorem 3.3.3 *A subset X of $w_1^* \cdots w_r^*$ is regular if and only if it is a finite union of sets of the form $X_1 X_2 \cdots X_r$ where each X_i is a regular subset of w_i^*.*

This is, implicitly, an equivalent statement to our Theorem 3.3.2, though the proof given by Ginsburg and Spanier [18] uses different methods. We continue with our proof of Theorem 3.3.2, found independently of the proof of Ginsburg and Spanier:

Proof: Clearly if

$$L = \bigcup_{i=1}^{N} w_1^{a_{i,1}}(w_1^{b_{i,1}})^* \cdots w_n^{a_{i,n}}(w_n^{b_{i,n}})^*$$

then L is regular.

We prove the converse by induction on n. For $n = 1$, let $L \subseteq w_1^*$. Let $L_0 \subseteq 0^*$ be the regular language such that $h(L_0) = L$, where $h(0) = w_1$. Then L_0 is a unary regular language, and by Theorem 3.2.1, $L_0 = A \cup B(0^n)^*$ for finite sets A and B. For appropriate choices of a_i and b_i, depending on A and B, we have $L_0 = \bigcup_{i=1}^{N} 0^{a_i}(0^{b_i})^*$. Then applying h gives $L = \bigcup_{i=1}^{N} w_1^{a_i}(w_1^{b_i})^*$.

Let n be any integer with $n > 1$. Assume that for any integer $k < n$, any words $w_1^*, \ldots, w_k^*$, and for any regular language $L \subseteq w_1^* \cdots w_k^*$, L can be expressed as

$$L = \bigcup_{i=1}^{N} w_1^{a_{i,1}}(w_1^{b_{i,1}})^* \cdots w_k^{a_{i,k}}(w_k^{b_{i,k}})^*$$

for appropriate constants $a_{i,j}$ and $b_{i,j}$.

Let L be a regular language with $L \subseteq w_1^* \cdots w_n^*$ and $M = (Q, \Sigma, \delta, q_0, F)$ be the minimal DFA with $L(M) = L$. Let $m = |Q|$ and $Q = \{q_0, \ldots, q_{m-1}\}$. We employ the Myhill-Nerode theorem (see [24, Thm 3.9, pp. 65–66]). For $0 \leq i \leq m-1$, let E_i be the equivalence class defined by the Myhill-Nerode relation:

$$x \equiv y \iff (\forall z)(xz \in L \iff xy \in L).$$

Define the sets E_i' for $0 \leq i \leq m-1$ as

$$E_i' = E_i \cap w_1^* \cdots w_{n-1}^*.$$

E_i' is regular, since E_i is accepted by the automaton $M_i = (Q, \Sigma, \delta, q_0, \{q_i\})$, if we identify E_i with the state q_i of the minimal DFA M.

We also define F_i for $0 \leq i \leq m-1$ as follows:

$$F_i = \{y \in w_n^* \ : \ \exists x \in E_i' \text{ such that } xy \in L\}.$$

From Lemma 3.3.1, each F_i is regular. We claim that for all i with $0 \leq i \leq m-1$

$$E_i' F_i \subseteq L.$$

To see this, let $x \in E_i'$ and $y \in F_i$. By definition of F_i, there exists $x_y \in E_i'$ such that $x_y y \in L$. But $x_y \equiv x$ by definition of E_i'. So $x_y y \in L$ implies $xy \in L$. Thus $E_i' F_i \subseteq L$. Note that $F_i = \emptyset$ is possible, which implies $E_i' F_i \subseteq L$ trivially.

We define a set $J \subseteq \{0, \ldots, m-1\}$ as follows:

$$j \in J \iff \exists k \text{ such that } \delta(q_j, w_n^k) \in F$$

Since $E_i' F_i \subseteq L$ for all $0 \leq i \leq m-1$,

$$\bigcup_{j \in J} E_j' F_j \subseteq L.$$

We claim the reverse inclusion also holds. Let $x \in L$. Then $\delta(q_0, x) = q_s$ for some $q_s \in F$. Let x_1 be the prefix of x which lies in $w_1^* \cdots w_{n-1}^*$. Let $\delta(q_0, x_1) = q_r$ for some r. By definition of M and E_r, $x_1 \in E_r$, and so $x_1 \in E_r'$. Further, if we write $x = x_1 x_2$, where x_2 lies in w_n^*, $\delta(q_r, x_2) = q_s \in F$, so that, by definition, $r \in J$.

Also, we have that $x_2 \in F_r$. Thus, $x = x_1 x_2 \in E'_r F_r \subseteq \bigcup_{j \in J} E'_j F_j$.

Thus, we conclude that

$$L = \bigcup_{j \in J} E'_j F_j$$

since $J \subseteq \{1, \ldots, m\}$ the union is finite. But now by the inductive hypothesis,

$$E'_j = \bigcup_{i=1}^{N_j} w_1^{a_{1,i,j}} (w_1^{b_{1,i,j}})^* \cdots w_{n-1}^{a_{n-1,i,j}} (w_{n-1}^{b_{n-1,i,j}})^*$$

$$F_j = \bigcup_{k=1}^{M_j} w_n^{a_{n,k,j}} (w_n^{b_{n,k,j}})^*$$

So that

$$L = \bigcup_{j \in J} \bigcup_{i=1}^{N_j} \bigcup_{k=1}^{M_j} w_1^{a_{1,i,j}} (w_1^{b_{1,i,j}})^* \cdots w_{n-1}^{a_{n-1,i,j}} (w_{n-1}^{b_{n-1,i,j}})^* w_n^{a_{n,k,j}} (w_n^{b_{n,k,j}})^*$$

□

We now use Theorem 3.3.2 as a means of classifying bounded regular languages with minimal regular covers.

Definition 3.3.4 *Let $L \subseteq w_1^* \cdots w_n^*$ be a bounded language. Given fixed integers $a_{i,j}$ and $b_{i,j}$ $(1 \leq i \leq N,\ 1 \leq j \leq n)$, we define, for each $1 \leq i \leq N$ and $1 \leq k \leq n$, the sets*

$$\begin{aligned} &I_{i,k}(L, j_1, \ldots, j_{k-1}, j_{k+1}, \ldots, j_n) \\ &\quad = \{j : w_1^{a_{i,1}+b_{i,1}j_1} \cdots w_{k-1}^{a_{i,k-1}+b_{i,k-1}j_{k-1}} w_k^{a_{k,i}+b_{k,i}j} w_{k+1}^{a_{i,k+1}+b_{i,k+1}j_{k+1}} \cdots w_n^{a_{i,n}+b_{i,n}j_n} \in L\} \end{aligned}$$

for any $j_1, j_2, \ldots, j_{i-1}, j_{i+1}, \ldots, j_n \geq 0$.

Theorem 3.3.5 *Let* $L \subseteq R \subseteq w_1^* \cdots w_n^*$ *be bounded languages. Let* R *be regular, with representation*

$$R = \bigcup_{i=1}^{N} w_1^{a_{i,1}} (w_1^{b_{i,1}})^* \cdots w_n^{a_{i,n}} (w_n^{b_{i,n}})^*.$$

Then L *has minimal regular cover* R *if and only if for all* $1 \leq i \leq N$, $1 \leq k \leq n$ *and for all* $j_1, j_2, \ldots, j_{i-1}, j_{i+1}, \ldots, j_n \geq 0$, *the set*

$$I_{i,k}(L, j_1, j_2, \ldots, j_{i-1}, j_{i+1}, \ldots, j_n)$$

intersects every arithmetic progression.

Proof: Let $L \subseteq R$. We write $R = \bigcup_{i=1}^{N} R_i$ where

$$R_i = w_1^{a_{i,1}} (w_1^{b_{i,1}})^* \cdots w_n^{a_{i,n}} (w_n^{b_{i,n}})^*.$$

Let $L_i = L \cap R_i$. Since minimal regular covers are closed under intersection with regular languages, R_i is a minimal regular cover of L_i if and only if R is a minimal regular cover L. Note that $L = \bigcup_{i=1}^{N} L_i$. Thus, it suffices to show the result for L_i, since minimal regular covers are closed under union.

First, assume that for L_i, there exist k and $j_1, j_2, \ldots, j_{k-1}, j_{k+1}, \ldots, j_n \geq 0$ such that

$$I = I_{i,k}(L_i, j_1, j_2, \ldots, j_{k-1}, j_{k+1}, \ldots, j_n)$$

does not intersect every arithmetic progression. Let a, b be integers with $a \neq 0$ such that $am + b \notin I$ for all $m \geq 0$. Define D as the set

$$D = w_1^{a_{i,1}+b_{i,1}\cdot j_1} \cdots w_{k-1}^{a_{i,k-1}+b_{i,k-1}\cdot j_{k-1}} w_k^{a_{k,i}+b_{k,i}\cdot a} (w_k^{a_{k,i}+b_{k,i}\cdot b})^* w_{k+1}^{a_{i,k+1}+b_{i,k+1}\cdot j_{k+1}} \cdots w_n^{a_{i,n}+b_{i,n}\cdot j_n}.$$

Clearly D is infinite and regular. By the choice of D, we claim that $D \cap L_i = \emptyset$ and $D \subseteq R_i$.

By definition of I, we know that if

$$w_1^{a_{i,1}+b_{i,1}\cdot j_1} \cdots w_{k-1}^{a_{i,k-1}+b_{i,k-1}\cdot j_{k-1}} w_k^{a_{k,i}+b_{k,i}\cdot a} (w_k^{a_{k,i}+b_{k,i}\cdot b})^m w_{k+1}^{a_{i,k+1}+b_{i,k+1}\cdot j_{k+1}} \cdots w_n^{a_{i,n}+b_{i,n}\cdot j_n} \in L_i,$$

there exists an m with $am + b \in I$, contrary to our assumption on a, b and I. So, we must conclude that $D \cap L_i = \emptyset$. Also, $D \subseteq R_i$, by the definition of D. So, define $R_i' = R_i - D$. Then $L_i \subseteq R_i' \subseteq R_i$ and $R_i - R_i' = D$ is infinite. So R_i is not a minimal regular cover of L_i.

To prove the converse, assume that L_i is not minimally covered by R_i. Then there exists an R_i' such that $L_i \subseteq R_i' \subseteq R_i$ and $R_i - R_i'$ is infinite. Recall that

$$R_i = w_1^{a_{i,1}}(w_1^{b_{i,1}})^* \cdots w_n^{a_{i,n}}(w_n^{b_{i,n}})^*.$$

Let $R_i - R_i'$ be written as

$$R_i - R_i' = \bigcup_{\ell=1}^{M} w_1^{c_{\ell,1}}(w_1^{d_{\ell,1}})^* \cdots w_n^{c_{\ell,n}}(w_n^{d_{\ell,n}})^*.$$

Since $R_i - R_i'$ is infinite, we must have that there is an ℓ, $1 \leq \ell \leq M$, such that

$$w_1^{c_{\ell,1}}(w_1^{d_{\ell,1}})^* \cdots w_n^{c_{\ell,n}}(w_n^{d_{\ell,n}})^* \tag{3.3.1}$$

is infinite. Further, there must be a k, $1 \leq k \leq n$ such that $d_{\ell,k} \neq 0$, since (3.3.1) is infinite. So, choose arbitrary $j_1, j_2, \ldots, j_{k-1}, j_{k+1}, \ldots, j_n$ such that

$$w_1^{c_{\ell,1}+d_{\ell,1}\cdot j_1} \cdots w_{k-1}^{c_{\ell,k-1}+d_{\ell,k-1}\cdot j_{k-1}} w_k^{c_{\ell,k}}(w_k^{d_{\ell,k}})^* w_{k+1}^{c_{\ell,k+1}+d_{\ell,k+1}\cdot j_{k+1}} \cdots w_n^{c_{\ell,n}+d_{\ell,n}\cdot j_n}$$

is infinite. Define D to be this subset of $R_i - R_i'$:

$$D = w_1^{c_{\ell,1}+d_{\ell,1}\cdot j_1} \cdots w_{k-1}^{c_{\ell,k-1}+d_{\ell,k-1}\cdot j_{k-1}} w_k^{c_{\ell,k}}(w_k^{d_{\ell,k}})^* w_{k+1}^{c_{\ell,k+1}+d_{\ell,k+1}\cdot j_{k+1}} \cdots w_n^{c_{\ell,n}+d_{\ell,n}\cdot j_n}.$$

Since $D \subseteq R_i$, each string is a member of R_i. In particular, if we define the strings U_m for each $m \geq 0$ as

$$U_m = w_1^{c_{\ell,1}+d_{\ell,1}\cdot j_1} \cdots w_{k-1}^{c_{\ell,k-1}+d_{\ell,k-1}\cdot j_{k-1}} w_k^{c_{\ell,k}+d_{\ell,k}\cdot m} w_{k+1}^{c_{\ell,k+1}+d_{\ell,k+1}\cdot j_{k+1}} \cdots w_n^{c_{\ell,n}+d_{\ell,n}\cdot j_n},$$

each $U_m \in R_i$. Consider first U_0. We want to write U_0 in terms of $a_{i,j}$ and $b_{i,j}$, which is possible since $U_0 \in R_i$. So, let $j_1', \cdots j_n'$ be integers such that

$$U_0 = w_1^{a_{i,1}+b_{i,1}\cdot j_1'} \cdots w_{k-1}^{a_{i,k-1}+b_{i,k-1}\cdot j_{k-1}'} w_k^{a_{i,k}+b_{i,k}\cdot j_k'} w_{k+1}^{a_{i,k+1}+b_{i,k+1}\cdot j_{k+1}'} \cdots w_n^{a_{i,n}+b_{i,n}\cdot j_n'}.$$

Equating powers in our two representations of U_0, we have that

$$c_{\ell,r} + d_{\ell,r} \cdot j_r = a_{i,r} + b_{i,r} \cdot j'_r$$

for $1 \leq r \leq n$, $r \neq k$, and

$$c_{\ell,k} = a_{i,k} + b_{i,k} \cdot j'_k.$$

Next, consider U_1. We know that U_0 and U_1 share the same prefix of length $\sum_{i=1}^{k-1}(c_{\ell,i} + d_{\ell,i} \cdot j_i)|w_i|$ (that is, the prefix of powers of all $w_1, \ldots, w_{k-1}$) and a suffix of length $\sum_{i=k+1}^{n}(c_{\ell,i} + d_{\ell,i} \cdot j_i)|w_i|$ (similarly, the suffix of powers of all $w_{k+1}, \ldots, w_n$). We can then use the same $j'_1, \ldots, j'_{k-1}, j'_{k+1}, \ldots, j'_n$ to write U_1 as

$$U_1 = w_1^{a_{i,1}+b_{i,1}\cdot j'_1} \cdots w_{k-1}^{a_{i,k-1}+b_{i,k-1}\cdot j'_{k-1}} w_k^{(a_{i,k}+b_{i,k}\cdot j'_k)+b_{i,k}\cdot c} w_{k+1}^{a_{i,k+1}+b_{i,k+1}\cdot j'_{k+1}} \cdots w_n^{a_{i,n}+b_{i,n}\cdot j'_n}$$

for some constant c. Similarly, by comparing lengths, we conclude that we may write

$$U_m = w_1^{a_{i,1}+b_{i,1}\cdot j'_1} \cdots w_{k-1}^{a_{i,k-1}+b_{i,k-1}\cdot j'_{k-1}} w_k^{(a_{i,k}+b_{i,k}\cdot j'_k)+b_{i,k}\cdot c\cdot m} w_{k+1}^{a_{i,k+1}+b_{i,k+1}\cdot j'_{k+1}} \cdots w_n^{a_{i,n}+b_{i,n}\cdot j'_n}.$$

Now, each string $U_m \in R_i - R'_i$ and thus, since $L_i \subseteq R'_i$, $U_m \notin L_i$ for all $m \geq 0$. Thus, by definition of

$$I = I_{i,k}(L_i, j'_1, \ldots, j'_{k-1}, j'_{k+1}, \ldots, j'_n)$$

we have that $j'_k + c \cdot m \notin I$ for all $m \geq 0$. This proves the theorem. $\square$

3.4 Decision problems for MRCs

We examine some decision problems for minimal regular covers. Our first result shows that it is undecidable whether an arbitrary CFL has a minimal regular cover. We show the same is true even if we restrict ourselves to linear CFLs.

Theorem 3.4.1 *Let L be an arbitrary CFL. The following are undecidable:*

(a). whether L has a minimal regular cover.

(b). whether $\overline{L}$ has a minimal regular cover.

We appeal to Greibach's Theorem [20]:

Theorem 3.4.2 *Let $\mathcal{C}$ be a class of languages that is effectively closed under concatenation with regular sets and union, and for which the decision problem "$= \Sigma^*$" is undecidable for any sufficiently large fixed Σ. Let P be a property that is true for all regular sets and that is preserved under quotient with a single letter. Then P is undecidable for $\mathcal{C}$.*

We now prove Theorem 3.4.1:

Proof: (a) Note that CFLs satisfy the conditions of Greibach's Theorem (see [24, pp. 205–206]). Define $P(L) = \{L$ has a minimal regular cover$\}$. By Theorem 3.1.5, P is closed under arbitrary quotient. Finally, we note that every regular language trivially has a minimal regular cover.

(b) Let $P'(L) = \{\overline{L}$ has a minimal regular cover$\}$. Again, P' is true for all regular sets. Assume that $P'(L)$ is true. Now, we note that $P'(L/a)$ is true if and only if $\overline{L/a} = \Sigma^* - L/a$ has a minimal regular cover. But it is easy to observe that $\Sigma^* = \Sigma^*/a$, and so that $\Sigma^* - L/a = \Sigma^*/a - L/a = (\Sigma^* - L)/a$. But by Theorem 3.1.5, we must have that $(\Sigma^* - L)/a$ has minimal regular cover, since we assume that $P'(L)$ is true. So $P'(L/a)$ is true and thus we have satisfied the conditions for Greibach's Theorem. □

Recall that a grammar $G = (V, \Sigma, P, S)$ is said to be linear if each production $\alpha \to \beta_1 \cdots \beta_n$ satisfies $\beta_i \in V$ for at most one index i. That is, for all productions, all but at most one of the items on the right-hand side is a variable, the rest being elements of Σ.

We note the theorem of Hunt and Rosenkrantz [26] which gives a condition for undecidability of problems relating to linear CFLs over $\{0, 1\}$. We begin with a definition:

Definition 3.4.3 *Let P be a nontrivial predicate on a class $\mathcal{C}$ of languages over $\{0, 1\}$. Let r, s, x, y be strings over $\{0, 1\}$. Define the languages L_α and L_β by*

$$L_\alpha = \{0x, 1y\}^+, L_\beta = \{x0, y1\}^+$$

and the homomorphisms $h_\alpha, h_\beta : \{0, 1\}^* \to \{0, 1\}^*$ by

$$h_\alpha(0) = 0x,\ \ h_\alpha(1) = 1y$$

$$h_\beta(0) = x0,\ \ h_\beta(1) = y1.$$

Then we define $L(P, r, s, x, y)$ and $R(P, r, s, x, y)$ by

- $L(P, r, s, x, y) = \{h_\alpha^{-1}(u \setminus (r \setminus L/s))\ :\ u \in L_\alpha, P(L)$ *is true*$\}$,

- $R(P, r, s, x, y) = \{h_\beta^{-1}((r \setminus L/s)/v) \;:\; v \in L_\beta, P(L) \textit{ is true}\}$.

We now present the theorem of Hunt and Rosenkrantz [26]:

Theorem 3.4.4 *Let P be any predicate on the linear CFLs over $\{0,1\}$ for which there exists a linear CFL L_t and strings r, s, x and y in $\{0,1\}^*$ such that*

1. *$P(L_t)$ is true,*
2. *$r\{0x, 1y\}^*s \subseteq L_t$, and*
3. *$L(P, r, s, x, y)$ is a proper subset of the linear CFLs over $\{0,1\}$.*

or symmetrically,

1. *$P(L_t)$ is true,*
2. *$r\{x0, y1\}^*s \subseteq L_t$, and*
3. *$R(P, r, s, x, y)$ is a proper subset of the linear CFLs over $\{0,1\}$.*

Then for arbitrary linear CFG G, the predicate $P(L(G))$ is undecidable.

We demonstrate that the conditions of Theorem 3.4.4 hold for the predicate $P(L) = \{L \text{ has a minimal regular cover}\}$.

Theorem 3.4.5 *Let G be an arbitrary linear CFG over $\{0,1\}$. Then it is undecidable whether $L(G)$ has a minimal regular cover.*

Proof: We first note that h_α is an ϵ-free homomorphism, so that minimal regular covers are preserved under h_α^{-1}, by Theorem 3.1.11. Note that any unbounded regular language R satisfies the first two conditions of Theorem 3.4.4. In particular, let $L_t = \{0,1\}^*$. Then $r = s = x = y = \epsilon$ satisfies $r\{0x, 1y\}^*s \subseteq L_t$. Observe that

$$\begin{aligned} \mathcal{R} &= R(P, \epsilon, \epsilon, \epsilon, \epsilon) \\ &= \{L/x \;:\; x \in \{0,1\}^+, P(L) \text{ is true}\}. \end{aligned}$$

We must show that $\mathcal{R}$ is a proper subset of the linear CFLs over $\{0,1\}$. But note that by Theorem 3.6.5, $L_0 = \{0^n1^n \ : \ n \geq 0\}$ is a linear CFL which has no minimal regular cover. But were $L_0 \in \mathcal{R}$, then $L_0 = L_1/x$ for some $x \in \{0,1\}^+$ and with $P(L_1)$ is true. But by Theorem 3.1.5, $P(L_1)$ true implies $P(L_1/x)$ is true. This contradicts that L_0 has no minimal regular cover. Thus, it must be that $\mathcal{R}$ is a proper subset of the linear CFLs, since $L_0 \notin \mathcal{R}$. So our predicate P satisfies Theorem 3.4.4. □

Consider the following technical lemma:

Lemma 3.4.6 *Let L be any language and let x be any string. Then $\overline{L}/x = \overline{(L/x)}$.*

Proof: By definition, $\overline{(L/x)} = \overline{\{y \ : \ yx \in L\}} = \{y \ : \ yx \notin L\}$. But also by definition, $\overline{L}/x = \{y \ : \ yx \in \overline{L}\}$. This proves the lemma. □

The following is a simple lemma:

Lemma 3.4.7 *Let $L_0 = \{0^n1^n \ : \ n \geq 0\}$. Then $\overline{L_0}$ is a linear CFL.*

Proof: Observe that

$$\begin{aligned} \overline{L_0} &= \{0^n1^m \ : \ m,n \geq 0, m \neq n\} \cup \overline{\{0^*1^*\}} \\ &= \{0^n1^m \ : \ m,n \geq 0, m \neq n\} \cup 0^*11^*0(0+1)^* \end{aligned}$$

The second term in this union is regular, and thus there is a right linear grammar $G_1 = (V_1, \{0,1\}, P_1, S_1)$ such that $L(G_1) = 0^*11^*0(0+1)^*$. Further, the language $\{0^n1^m \ : \ m,n \geq 0, m \neq n\}$ can be given by the grammar $G_2 = (V_2, \{0,1\}, P_2, S_2)$

such that P_2 is given by

$$S_2 \to 0S_m|S_l1 \quad S_m \to 0S_m|S_s$$

$$S_l \to S_l1|S_s \quad S_s \to 0S_s1|\epsilon.$$

This is easily verified. So, we can combine G_1 and G_2 to give a linear grammar for $\overline{L_0}$. □

With this, we can prove the following:

Theorem 3.4.8 *Let G be an arbitrary linear CFG over $\{0,1\}$. Then it is undecidable whether $\overline{L(G)}$ has a minimal regular cover.*

Proof: The proof uses the same construction as the proof of Theorem 3.4.5, with the predicate $P(L) = \{\overline{L}$ has a minimal regular cover$\}$. Again, $L_t = \{0,1\}^*$, and $r = s = x = y = \epsilon$ satisfies the first two conditions of Theorem 3.4.4, since $\overline{L_t} = \emptyset$ has itself as minimal regular cover. So, again we have

$$\mathcal{R} = \{L/x \ : \ x \in \{0,1\}^+, P(L) \text{ is true}\}.$$

Now, we consider $L_0 = \{0^n1^n \ : \ n \geq 0\}$. By our Lemma 3.4.7, $\overline{L_0}$ is a linear CFL. Now, assume that $\overline{L_0} \in \mathcal{R}$. So, there must exist an $x \in \{0,1\}^+$ and a language L_1 such that $\overline{L_0} = L_1/x$, and $P(L_1)$ is true. But by Lemma 3.4.6,

$$L_0 = \overline{(L_1/x)} = \overline{L_1}/x \tag{3.4.1}$$

Since $P(L_1)$ is true, $\overline{L_1}$ has a minimal regular cover. Then by Theorem 3.1.5 and (3.4.1), L_0 has a minimal regular cover. But by Theorem 3.6.5, L_0 has no minimal regular cover. This is a contradiction and $\mathcal{R}$ is a proper subset of the linear CFLs.

□

3.5 Concerning bounded CFLs and MRCs

In Section 3.4, we saw that the problem of deciding whether a given CFL or linear CFL has a minimal regular cover is undecidable. In this section, we consider the equivalent problem for bounded CFLs. Results of Hunt, Rosenkrantz and Szymanski will allow us to conclude that this and related problems are hard. We have the following theorem of Hunt, Rosenkrantz and Szymanski [28, Theorem 4.11, p. 253]:

Theorem 3.5.1 *Let P be any nontrivial predicate on the bounded CFLs such that P is true for all bounded regular sets and either*

$$\mathcal{L} = \{L' \ : \ L' = x \setminus L, x \in \{0,1\}^*, P(L) \text{ is true } \}$$

or

$$\mathcal{R} = \{L' \ : \ L' = L/x, x \in \{0,1\}^*, P(L) \text{ is true } \}$$

is a proper subset of the bounded CFLs. Then there exists a constant $c > 0$ such that any nondeterministic multi-tape TM that recognizes

$$\{G \ : \ G \text{ is a CFG generating a bounded language and } P(L(G)) \text{ is false}\}$$

requires time at least $2^{cn/(\log n)}$ *infinitely often.*

Note that $\{0^n 1^n \ : \ n \geq 0\}$ is a bounded CFL with no minimal regular cover and that L/x is a bounded CFL for any bounded CFL L. Thus, in order for the the predicate $P(L) = \{L \text{ has a minimal regular cover}\}$ to satisfy the conditions of Theorem 3.5.1, we need only satisfy the condition that there exists a language L' such that $L' \neq L/x$ for all L bounded CFLs with $P(L)$ true. But by Theorem 3.1.5, if $P(L)$ is true, then $P(L')$ must also be true. Since we know that P is nontrivial on the bounded CFLs, we must conclude that an $\mathcal{R}$ is a proper subset of the bounded CFLs.

Thus we have the following theorem:

Theorem 3.5.2 *There exists a constant* $c > 0$ *such that any multi-tape NTM which recognizes*

$$\{G \ : \ G \text{ is a CFG with } L(G) \text{ bounded and } L(G) \text{ has no minimal regular cover}\}$$

requires time at least $2^{cn/(\log n)}$ *infinitely often.*

Given Lemma 3.4.7, the following theorem is proved similarly using the predicate $P(L) = \{\overline{L} \text{ has a minimal regular cover }\}$:

Theorem 3.5.3 *There exists a constant* $c > 0$ *such that any multi-tape NTM which recognizes*

$$\{G \ : \ G \text{ is a CFG with } L(G) \text{ bounded and } \overline{L(G)} \text{ has no minimal regular cover}\}$$

requires time at least $2^{cn/(\log n)}$ *infinitely often.*

Analogously, the theorem of Hunt *et al.* [28, Theorem 4.12, p. 224] which states:

Theorem 3.5.4 *Let* P *be any nontrivial predicate on the bounded linear CFLs such that* P *is true for all bounded regular sets and either*

$$\mathcal{L} = \{L' \ : \ L' = x \setminus L, x \in \{0,1\}^*, P(L) \text{ is true } \}$$

or

$$\mathcal{R} = \{L' \ : \ L' = L/x, x \in \{0,1\}^*, P(L) \text{ is true } \}$$

is a proper subset of the bounded linear CFLs. Then

$$\{G \ : \ G \text{ is a linear CFG generating a bounded language and } P(L(G)) \text{ is false}\}$$

is **NP**-*hard.*

gives us the following result:

Theorem 3.5.5 *The following problems are* **NP**-*hard:*

$L_1 = \{G \ : \ G$ *is a linear CFG,* $L(G)$ *is bounded and has no min. regular cover*$\}$,

$L_2 = \{G \ : \ G$ *is a linear CFG,* $\overline{L(G)}$ *is bounded and has no min. regular cover*$\}$.

3.6 Examples

In this section, we consider some particular classes of languages and their minimal regular covers. In particular we demonstrate that there exist context free languages which do not have a minimal regular cover. We begin with the theorems of Lyndon and Schützenberger [37]:

Theorem 3.6.1 *Let* $y \in \Sigma^*$ *and* $x, z \in \Sigma^+$. *Then* $xy = yz$ *if and only if there exist* $u, v \in \Sigma^*$ *and an integer* $e \geq 0$ *such that* $x = uv$, $z = vu$ *and* $y = (uv)^e u = u(vu)^e$.

Theorem 3.6.2 *Let* $x, y \in \Sigma^+$. *Then the following three conditions are equivalent:*

- $xy = yx$;
- *There exist integers* $i, j > 0$ *such that* $x^i = y^j$;
- *There exist* $z \in \Sigma^+$ *and integers* $k, l > 0$ *such that* $x = z^k$ *and* $y = z^l$.

We also have the theorem of Fine and Wilf [12]:

Theorem 3.6.3 *Let* $x, y \in \Sigma^*$. *If there exist integers* $p, q \geq 0$ *such that* x^p *and* y^q *have a common prefix (resp., suffix) of length at least* $|x| + |y| - \gcd(|x|, |y|)$, *then there exists* $z \in \Sigma^*$ *and integers* $i, j \geq 0$ *such that* $x = z^i$ *and* $y = z^j$.

The following lemma is due to Shallit and Yu:

Lemma 3.6.4 *Let* $u, v, w, x, y \in \Sigma^*$ *be fixed words with* $v, x \neq \epsilon$. *Define* $L = L(u, v, w, x, y) = \{uv^i wx^i y \ : \ i \geq 0\}$. *Then the following conditions are equivalent:*

(a). L *is regular;*

(b). L *has an infinite regular subset;*

(c). There exist integers $k, l \geq 1$ *such that* $v^k w = wx^l$;

(d). There exist words $r, s \in \Sigma^*$ *and integers* $m, n \geq 1$ *and* $p \geq 0$ *such that* $v = (rs)^m$, $w = (rs)^p r$ *and* $x = (sr)^n$;

(e). There exist integers $a, b \geq 0$ *with* $a \neq b$ *such that* $uv^a wx^b y \in L$.

This result was found independently of Hunt *et al.* [28, pp. 239–240], who proved the implications (e) $\iff$ (b) and (e) $\Rightarrow$ (a). The equivalence of (a) and (c) was also proved by Latteux and Therrin [31, Lemma 5.5, p. 14].

Proof: We prove (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (a), and (c) $\iff$ (e).

(a) $\Rightarrow$ (b): L is clearly infinite, so if L is regular, then it has an infinite regular subset, namely L itself.

(b) $\Rightarrow$ (c): Suppose R is an infinite regular language, and $R \subseteq L$. Then by the pumping lemma, there exist words r, $s \neq \epsilon$, and t such that $rs^*t \subseteq R$. Then there exist integers $m, n \geq 0$ such that $rt = uv^m wx^m y$ and $rst = uv^{m+n} wx^{m+n} y$. Since L contains at most one word of each length, it then follows that $rs^it = uv^{m+ni} wx^{m+ni} y$ for all $i \geq 0$.

By replacing r with rs^a for some a and t with s^bt for some b, if necessary, we may assume without loss of generality that $|r| > |u|$ and $|t| > |y|$. Similarly, by replacing s with s^c for some c, if necessary, we may assume $|r| < |uv^{m+ni}|$ and $|t| < |x^{m+ni}y|$ for all $i \geq 1$.

It now follows that $r = uv^e v'$, $s = v''v^f wx^g x'$, $t = x''x^h y$ for some $e, f, g, h \geq 0$ where $v = v'v''$ and $x = x'x''$. Now choose i sufficiently large such that $m + ni - e - h - 1 > |s| + |v| + |x|$. Then

$$rs^it = uv^{m+ni} wx^{m+ni} y.$$

By cancelling $r = uv^e v'$ on the left and $t = x''x^h y$ on the right, we get

$$\begin{aligned} s^i &= v''v^{m+ni-e-1}wx^{m+ni-h-1}x' \\ &= (v''v')^{m+ni-e-1}v''wx'(x''x')^{m+ni-h-1}. \end{aligned} \tag{3.6.1}$$

We now observe that s^i and $(v''v')^{m+ni-e-1}$ agree on a prefix of size $\geq m+ni-e-1 > |s| + |v|$, and so by Theorem 3.6.3, there exists a word Y such that s and $v''v'$ are both powers of Y. Similarly, s^i and $(x''x')^{m+ni-h-1}$ agree on a suffix of size $\geq m + ni - h - 1 > |s| + |x|$, and so by Theorem 3.6.3, there exists a word Z such that s and $x''x'$ are both powers of Z. From Theorem 3.6.2 it now follows that there exists a word X such that both Y and Z are powers of X. Hence s, $v''v'$, and $x''x'$ are all powers of X. Then from Eq. (3.6.1) it follows that $v''wx'$ is a power of X. Now write $v''v' = X^l$, $x''x' = X^k$, and $v''wx' = X^d$ for some integers l, k, d. Then

$$(v''v')^k v''wx' = v''wx'(x''x')^l.$$

Now cancel v'' on the left and x' on the right; we get

$$(v'v'')^k w = w(x'x'')^l,$$

and hence $v^k w = wx^l$, as desired.

(c) $\Rightarrow$ (d): If $v^k w = wx^l$, then by Theorem 3.6.1 there exist $t, z \in \Sigma^*$ and an integer $e \geq 0$ such that $v^k = tz$, $w = t(zt)^e$, and $x^l = zt$.

Now $v^k = tz$ implies there exists a decomposition $v = v'v''$ such that $t = v^i v'$ and $z = v''v^j$ for some integers $i, j \geq 0$. Then $x^l = zt = v''v^j v^i v' = (v''v')^{i+j+1}$. Set $h = v''v'$. Then $x^l = h^{i+j+1}$. Hence, by Theorem 3.6.2, there exists a word $g \in \Sigma^+$ and integers $n, a \geq 1$ such that $x = g^n$ and $h = g^a$.

From this last equality we get $v''v' = g^a$. Thus there exists a decomposition $g = sr$ such that $v'' = g^b s$ and $v' = rg^c$ for some integers $b, c \geq 0$. Then we find $v = v'v'' = (rg^c)(g^b s) = (rs)^{b+c+1}$.

Finally, we have

$$\begin{aligned} w &= t(zt)^e \\ &= v^i v'(x^l)^e \\ &= (rs)^{(b+c+1)i} rg^c (g^n)^{le} \\ &= (rs)^{(b+c+1)i} r(sr)^{c+nle} \\ &= (rs)^{(b+c+1)i+c+nle} r. \end{aligned}$$

so (d) holds with $m = b + c + 1$, and $p = (b + c + 1)i + c + nle$.

(d) $\Rightarrow$ (a): For all $i \geq 0$ we have

$$uv^i wx^i y = u(rs)^{mi}(rs)^p r(sr)^{ni} y = u(rs)^{(m+n)i+p} ry$$

so $L = u((rs)^{m+n})^*(rs)^p ry$, which is clearly regular. $\square$

Theorems 3.6.5 and 3.6.6 are due to Shallit and Yu.

Theorem 3.6.5 *Let $u, v, w, x, y \in \Sigma^*$ be fixed words with $v, x \neq \epsilon$. Let $L = \{uv^iwx^iy \ : \ i \geq 0\}$. Then L is context-free, and L has a minimal regular cover if and only if L is regular.*

Proof: It is clear that L is context-free, since it is generated by the following context-free grammar:

$$\begin{aligned} S &\rightarrow uBy \\ B &\rightarrow w \mid vBx \end{aligned}$$

First, suppose that L is minimally covered by the regular language R. Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA accepting R. Then there exists a state q such that $\delta(q_0, uv^iw) = q$ for infinitely many i. Let $I = \{i \ : \ \delta(q_0, uv^iw) = q\}$ and i_0 be the smallest element of I. Define

$$R' = \{x^jy \ : \ \delta(q, x^jy) \in F \text{ and } j > i_0\}.$$

Then R' is infinite since it contains $\{x^iy \ : \ i \in I - i_0\}$. Furthermore, R' is regular since

$$R' = \{t \in \Sigma^* \ : \ \delta(q, t) \in F\} \cap x^{i_0}x^+y.$$

Let $R'' = uv^{i_0}wR'$. Then R'' is infinite and regular, and clearly $R'' \subseteq R$. Now if $R'' \subseteq R - L$, then $R''' = R - R''$ is a regular cover of L with $R - R'''$ infinite. Hence R is not minimal, a contradiction. Thus $R'' \not\subseteq R - L$, and hence $uv^{i_0}wx^{j_0}y \in L$ for some $j_0 > i_0$. Then by Lemma 3.6.4, L is regular.

On the other hand, if L is regular then L itself is trivially a minimal regular cover. □

Theorem 3.6.6 *Let $u, v, w, x, y \in \Sigma^*$ be fixed words with $v, x \neq \epsilon$. Let $L = \{uv^iwx^iy \ : \ i \geq 0\}$. Then $\overline{L}$ is a context-free language, and further, if L is not regular, then Σ^* is a minimal regular cover of $\overline{L}$ with respect to finite languages.*

Proof: To see that $\overline{L}$ is context-free, we describe a pushdown automaton (PDA) M accepting L.

The PDA M accepts its input z in either one of two cases:

(a). The length $|z|$ is not of the form $|uwy| + i|vx|$ for some $i \geq 0$.

(b). The length $|z|$ is of the form $|uwy| + i|vx|$ for some $i \geq 0$, and a mismatch between z and uv^iwx^iy has been found somewhere, and $|uwy| + i|vx|$ input symbols have been processed.

Case (b) is the only tricky part. The PDA starts by comparing the first $|u|$ symbols of its input with the symbols of u. If a mismatch is detected, the PDA records this fact, but does not accept yet. Then it compares the next symbols of its input with v^i, where i is chosen nondeterministically. If a mismatch is detected, the PDA records this fact, but does not accept yet. For each copy of v compared to the input, a counter is placed in the stack. Eventually the PDA chooses (nondeterministically) to stop matching v's. When it does so, there are i counters on the stack. Now the PDA continues comparing the next $|w|$ symbols of its input with w. Again, if a mismatch is detected, it records this fact, but does not accept yet.

Then, using the count recorded in the stack, it compares the next $i|x|$ symbols of the input with x^i. Finally, it compares the next $|y|$ symbols of its input with y. When the computation is complete, the PDA has read $|uwy| + i|vx|$ symbols, and it accepts if and only if there was a mismatch somewhere along the way.

Now suppose L is not regular. Then $\overline{L}$ is not regular. We show Σ^* is a minimal regular cover of $\overline{L}$. Suppose not. Then there exists a regular language L' such that $\overline{L} \subseteq L' \subseteq \Sigma^*$, and $\Sigma^* - L'$ is infinite. But then $\Sigma^* - L' \subseteq L$, so L has an infinite regular subset. But then L is regular by Lemma 3.6.4, a contradiction. □

We now present an interesting class of languages which have minimal regular cover Σ^*. Let $0 < \alpha < 1$ be an irrational real number. Define

$$f_n = \lfloor (n+1)\alpha \rfloor - \lfloor n\alpha \rfloor.$$

We set $\mathbf{f}_\alpha = f_1 f_2 f_3 \cdots$, an infinite word over the alphabet $\{0, 1\}$. The infinite word $\mathbf{f}_\alpha$ is called the characteristic word of α. With each such α we associate the language $L_\alpha = \{0^i \ : \ f_i = 1\}$.

Theorem 3.6.7 *For all irrational real α with $0 < \alpha < 1$, the language L_α is not regular and is minimally covered by 0^*.*

Proof: If L_α were regular, then $\mathbf{f}_\alpha$ would be ultimately periodic. But then $\lim_{n\to\infty} \frac{\sum_{1 \le i \le n} f_i}{n} = \alpha$ would be rational, a contradiction.

To see that L_α is minimally covered by 0^*, suppose it is minimally covered by a regular language R with $0^* - R$ infinite. Then by the pumping lemma there exist words $u = 0^j, v = 0^k, w = 0^l$ such that $uv^iw \in 0^* - R$ for all $i \ge 0$. But

$uv^iw = 0^{j+l+ik}$. It now suffices to show that for all c, d with $c \geq 0, d \geq 1$, there exists an $i \geq 0$ such that $f_{c+id} = 1$. This easily follows from Kronecker's theorem [21, Theorem 438]. □

3.7 Bi-MRCs

We briefly consider the minimal cover analogue of bi-immunity [3]. We propose the following definition:

Definition 3.7.1 *A language L has bi-minimal regular cover (L_1, L_2) if L_1 is a minimal regular cover of L and L_2 is a minimal regular cover of $\overline{L}$.*

In the terminology of Balcázar and Schöning, a language $L \subseteq \Sigma^*$ has bi-minimal regular cover (Σ^*, Σ^*) if and only if L is **REG**-bi-immune. We will consider some classes of languages which have bi-minimal regular covers. We begin with an easy class of languages:

Example 3.7.2: Let $L_0 = \{0^{n+n!} \ : \ n \geq 0\}$.

Then by Lemma 3.2.2, $\overline{L_0}$ has minimal regular cover 0^*. By Lemma 3.3.5, L_0 has minimal regular cover 0^* if and only if every arithmetic progression intersects $I = \{n + n! \ : \ n \geq 0\}$. Let a, b be arbitrary integers with $a \neq 0$. Consider $m = \frac{(a+b)!}{a} + 1$. Then $am + b = a(\frac{(a+b)!}{a} + 1) + b = (a + b)! + a + b$. Thus for this choice of m, $am + b \in I$. Thus L_0 has bi-minimal regular cover $(0^*, 0^*)$.

By extension, let k be a positive integer. Let $L_k = \{0^{k(n+n!)} \ : \ n \geq 0\}$. Then $L_k = \varphi_k(L_0)$ where $\varphi_k(0) = 0^k$, so L_k has minimal regular cover $(0^k)^*$.

Now, observe that

$$\overline{L_k} = \{0^r \ : \ r \not\equiv 0 (\text{mod } k)\} \cup \{0^{kn} \ : \ n \neq m + m! \ \forall m \geq 0\} \tag{3.7.1}$$

The second term of the union in (3.7.1) is $\varphi_k(\overline{L_0})$, and thus has minimal regular cover $(0^k)^*$. But the first term of the union in (3.7.1) is regular, and so has minimal regular cover itself. Thus, $\overline{L_k}$ has minimal regular cover

$$\{0^r \ : \ r \not\equiv 0 (\text{mod } k)\} \cup (0^k)^* = 0^*.$$

Thus L_k has bi-minimal regular cover $((0^k)^*, 0^*)$. □

The class of languages related to characteristic words introduced in Section 3.6 gives us another class of languages which have bi-minimal regular covers. The following preliminary lemma is adapted from Lothaire [33, Cor. 2.2.20]:

Lemma 3.7.3 *Let α be an irrational number with $0 < \alpha < 1$. Then $L_{1-\alpha} = 0^* - L_\alpha$.*

Clearly α is irrational if and only if $1 - \alpha$ is irrational. Thus, Theorem 3.6.7 immediately gives us the following result:

Theorem 3.7.4 *Let α be an irrational number with $0 < \alpha < 1$. Then L_α has bi-minimal regular cover $(0^*, 0^*)$.*

Thus, by Theorem 2.3.6, for each irrational α with $0 < \alpha < 1$, L_α is not a DCFL. Since the CFLs and thus the DCFLs are precisely the regular languages over a unary alphabet, we have that each L_α is at least a CSL. We now demonstrate a large class of irrational numbers α which yield L_α a CSL. We use the theory of Sturmian words (see Allouche and Shallit [1, Ch. 9] or Lothaire [33, Ch. 2]).

Recall the definition of continued fraction. Every irrational number α admits a unique expansion of the form

$$\alpha = a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\cdots}}}$$

We denote this as

$$\alpha = [a_0, a_1, a_2, \ldots].$$

If $\alpha = [a_0, a_1, \ldots]$, we use the notation

$$\alpha = [a_0, a_1, a_2, \ldots, a_i + \beta]$$

to denote that $\beta = [0, a_{i+1}, a_{i+2}, \ldots]$. We define the homomorphism $h_m : \{0,1\}^* \to \{0,1\}^*$ for each $m \geq 1$: $h_m(0) = 0^{m-1}1$, $h_m(1) = 0^{m-1}10$. From the theory of Sturmian words, we have the following theorem about continued fractions [33, Cor. 2.2.22]:

Theorem 3.7.5 *If $\alpha = [0, a_1, a_2, \ldots, a_i, \beta]$ for some irrational number α with $0 < \alpha, \beta < 1$, then*

$$f_\alpha = (h_{a_1} \circ h_{a_2} \circ h_{a_3} \circ \cdots \circ h_{a_i})(f_\beta).$$

Call a continued fraction $[a_1, a_2, \ldots, a_i, \ldots]$ periodic if $\{a_i\}_{i\geq 1}$ is periodic as a sequence. Then we have the following corollary to Theorem 3.7.5 [1, Cor. 9.2.6]:

Corollary 3.7.6 *If* $\alpha = [0, a_1, a_2, \ldots]$ *has an ultimately periodic continued fraction expansion, then* f_α *is the fixed point of a homomorphism.*

We recall a remarkable theorem of Berstel [5] on the fixed point of a homomorphism:

Theorem 3.7.7 *Let* $h : \Sigma^* \to \Sigma^*$ *be a morphism that generates an infinite word on a letter* $a \in \Sigma$. *Let* F *be the set of left factors of this infinite word. Then* $\overline{F}$ *is a CFL.*

With Theorems 3.7.5 and 3.7.7, we can prove the following:

Theorem 3.7.8 *Let* α *be an irrational number with* $0 < \alpha < 1$. *If* α *has a ultimately periodic continued fraction then* L_α *is a CSL.*

Proof: By Corollary 3.7.6, f_α is a fixed point of a homomorphism h. Thus, let F_α be the set of non-empty left factors of f_α. By Theorem 3.7.7, $\overline{F_\alpha}$ is a CFL and so F_α is a CSL, since equation (1.1.1) gives that **CF** $\subseteq$ **CS** and it is known [29, 46] that **CS** = co-**CS**.

Now, consider $F' = F_\alpha \cap (0+1)^*1$. Since the CSLs are closed under intersection with regular languages (for the closure properties of the CSLs, see [24]), F' is still a CSL. But F' contains precisely those strings which end in a 1. That is, $x \in F' \iff f_{|x|} = 1$. Since the CSLs are closed under ϵ-free homomorphism, we conclude that $h(F') \subseteq 0^*$ is a CSL, where $h(0) = h(1) = 0$. But now $0^i \in h(F') \iff f_i = 1$. So $h(F') = L_\alpha$. Thus, L_α is a CSL. □

Thus, we have established unary **REG**-bi-immune languages which are CSLs. We can characterize the irrational numbers which give periodic continued fractions completely with the theorem of Lagrange (see [21, Thms 176 and 177]). Recall that an irrational number is quadratic if it is the irrational root of a quadratic equation with integer coefficients.

Theorem 3.7.9 *An irrational number has periodic continued fraction if and only if it is a quadratic irrational.*

Thus, not only do irrational numbers exist which yield CSLs, but we can restate our Theorem 3.7.8 more strongly as follows:

Theorem 3.7.10 *If α is a quadratic irrational number with $0 < \alpha < 1$, then L_α is a CSL.*

This was found independently of the results of Blanchard and Kòrka [6]. It is not clear whether the methods used by Blanchard and Kòrka can be used in this context.

Chapter 4

Minimal Context-Free Covers

The question of minimal context-free covers is covered in this chapter. We begin with a result showing that the languages with minimal context-free cover Σ^* are dense, due to Shallit:

Theorem 4.0.1 *Almost all languages $L \subseteq \Sigma^*$ have minimal context-free cover Σ^*.*

Proof: Let L' be a fixed context-free language with $\Sigma^* - L'$ infinite, say $\Sigma^* - L' = \{x_1, x_2, \ldots\}$, and suppose L' is a minimal cover of L. Then none of $x_1, x_2, \ldots$ are in L, which happens with measure 0. Now take the union over all context-free languages with $\Sigma^* - L'$ infinite; since this union is countable, the result has measure 0, too. □

The chapter begins with results on the closure properties of minimal context-free covers. We then examine a restricted case of minimal context-free covers of bounded languages, that is, the case when our language L is bounded by two words: $L \subseteq w_1^* w_2^*$. We then turn to a method for constructing languages with minimal context-free covers, by converting a regular language R to a context-free language $\{xx^R : x \in R\}$. The chapter ends with a consideration of languages with minimal context-free covers, and their commutative images under the Parikh mapping.

4.1 Closure properties of MCFCs

We examine some closure properties of minimal context free covers. We begin with a transitive property of minimal covers.

Theorem 4.1.1 *Let L_0 have minimal context-free cover L_1 and L_1 have minimal regular cover L_2. Then L_0 has minimal regular cover L_2.*

Proof: Assume, on the contrary, that L_3 is regular such that

$$L_0 \subseteq L_3 \subseteq L_2$$

and that $L_2 - L_3$ is infinite. Consider $L_3 \cap L_1$. We have that $L_0 \subseteq L_3 \cap L_1 \subseteq L_1$. Further, L_3 regular and L_1 context-free implies that $L_3 \cap L_1$ is context-free. But L_1 is a minimal context-free cover of L_0 so we must have that $L_1 - (L_3 \cap L_1) = L_1 - L_3$ is finite.

We have $L_3 \cup L_1$ is regular, since L_3 is regular and $L_1 - L_3$ is finite. Note that

$$L_1 \subseteq L_3 \cup L_1 \subseteq L_2$$

We have that $L_2 \neq L_1 \cup L_3$ since $L_2 - L_3$ is infinite, by assumption. We must have that $L_2 - (L_3 \cup L_1)$ is finite. This contradicts that $L_2 - L_3$ is infinite. We conclude that L_0 has minimal regular cover L_2. □

Recall that a if a family of languages is closed under intersection with regular sets, homomorphism and inverse homomorphism, it is termed a **full trio**. We note the following theorem [24, pp. 276–277], due to Ginsburg and Greibach [16]:

Theorem 4.1.2 *Every full trio is closed under quotient with a regular set.*

Let Σ be an alphabet and define $\hat{\Sigma} = \{\hat{a} \;:\; a \in \Sigma\}$. From the proof of Theorem 4.1.2 in Hopcroft and Ullman [24], we are given the following formulation:

$$L/R = h_2(h_1^{-1}(L) \cap (\hat{\Sigma})^* R) \tag{4.1.1}$$

where $h_1, h_2 : (\Sigma \cup \hat{\Sigma})^* \to \Sigma^*$ are homomorphisms defined by

$$h_1(a) = h_1(\hat{a}) = a$$

$$h_2(a) = \epsilon,\;\; h_2(\hat{a}) = a$$

Thus, we are able to make the following corollary

Corollary 4.1.3 *Let $\mathcal{C}$ be a full trio. If $\mathcal{C}$-minimal covers are closed under homomorphism, inverse ϵ-free homomorphism and intersection with regular languages, then $\mathcal{C}$-minimal covers are closed under quotient with regular languages as well.*

Proof: Let L_1 have minimal $\mathcal{C}$ cover L_2, and let R be a regular language. By Theorem 4.1.2, L_2/R is a member of $\mathcal{C}$. Note that h_1 is ϵ-free. Then by assumption, $h_1^{-1}(L_1)$ has minimal $\mathcal{C}$ cover $h_1^{-1}(L_2)$, $h_1^{-1}((L_1) \cap (\hat{\Sigma})^* R)$ has minimal $\mathcal{C}$ cover

$h_1^{-1}((L_2)\cap(\hat{\Sigma})^*R)$, and finally $h_2(h_1^{-1}(L_1)\cap(\hat{\Sigma})^*R)$ has minimal $\mathcal{C}$ cover $h_2(h_1^{-1}(L_2)\cap(\hat{\Sigma})^*R)$. Then by (4.1.1), L_1/R has minimal regular cover L_2/R. $\square$

4.2 Bounded MCFCs

In this section we examine minimal context-free covers as they result to bounded CFLs. We first recall two crucial results due to Ginsburg and Spanier [17], taken from Ginsburg [14, pp. 180–182]:

Theorem 4.2.1 *Let w_1, w_2 be words. If $L_1 \subseteq w_1^* w_2^*$ and L_2 are CFLs, then $L_1 \cap L_2$ is a CFL.*

Theorem 4.2.2 *Let w_1, w_2 be words. If $L_1 \subseteq w_1^* w_2^*$ and L_2 are CFLs, then $L_1 - L_2$ and $L_2 - L_1$ are CFLs.*

We can now prove that minimal context-free covers are unique up to finite modification, in the restricted case that one of the minimal context-free covers is bounded by two words:

Theorem 4.2.3 *Let L be a language, and let L_1, L_2 be minimal context-free covers of L, one of which is bounded by $w_1^* w_2^*$ for some words w_1, w_2. Then $(L_1 - L_2) \cup (L_2 - L_1)$ is finite.*

Proof: Assume without loss of generality that L_1 is bounded, say $L_1 \subseteq w_1^* w_2^*$ for words w_1, w_2.

Let $L_0 = L_1 \cap L_2$. By Theorem 4.2.1, L_0 is a context-free language with $L \subseteq L_0$ and thus $L_1 - L_0$ and $L_2 - L_0$ must both be finite by the minimality of L_1 and L_2. The result follows since $(L_1 - L_2) \cup (L_2 - L_1) = (L_1 - L_0) \cup (L_2 - L_0)$ □

We may also adapt Theorems 3.1.4 and 3.1.1 as follows:

Theorem 4.2.4 *Let L_1, L_2 be languages with minimal context-free covers L_1', L_2', respectively. If there exist words w_1, w_2 such that $L_1', L_2' \subseteq w_1^* w_2^*$ then $L_1 \cup L_2$ has minimal context-free cover $L_1' \cup L_2'$.*

Proof: Suppose L_3 is a context-free language such that $L_1 \cup L_2 \subseteq L_3 \subseteq L_1' \cup L_2'$. Note that $L_1' \cup L_2' \subseteq w_1^* w_2^*$. If $(L_1' \cup L_2') - L_3$ is infinite, then either $L_1' - L_3$ or $L_2' - L_3$ is infinite. Without loss of generality, assume that $L_1' - L_3 = L_1' - (L_1' \cap L_3)$ is infinite. But $L_1' \cap L_3$ satisfies $L_1 \subseteq L_1' \cap L_3 \subseteq L_1'$, and by Theorem 4.2.1, $L_1' \cap L_3$ is a CFL. This contradicts the minimality of L_1' as a context-free cover of L_1. □

Theorem 4.2.5 *Let L_1, L_2 be context-free languages with $L_2 \subseteq w_1^* w_2^*$ bounded. Then*

1. *If L is a language with minimal context-free cover L_1, then $L \cap L_2$ has minimal context-free cover $L_1 \cap L_2$.*

2. *If L is a language with minimal context-free cover L_2, then $L \cap L_1$ has minimal context-free cover $L_2 \cap L_1$.*

Proof: 1. Let L_3 be a regular language satisfying $L \subseteq L_3 \subseteq L_1 \cap L_2$. and $(L_1 \cap L_2) - L_3$ infinite. Note that $L_1 \cap L_2 \subseteq L_1 \subseteq w_1^* w_2^*$ and so it is bounded, and it is a CFL by Theorem 4.2.1. Then by an application of Theorem 4.2.2, we have that $(L_1 \cap L_2) - L_3 \subseteq L_1 \subseteq w_1^* w_2^*$ is a bounded CFL. So

$$L_0 = L_1 - ((L_1 \cap L_2) - L_3)$$

is also a CFL, by another application of Theorem 4.2.2. Now, $L \subseteq L_0 \subseteq L_1$. Also,

$$L_1 - L_0 = (L_1 \cap L_2) - L_3$$

the latter we assumed was infinite. This contradicts the minimality of L_1 as a context-free cover of L.

Item 2 follows by symmetry. □

We now characterize languages which have a minimal context-free cover, when the language is bounded by two words. We use the notation established in Ginsburg [14, p.150]:

Definition 4.2.6 *For subsets $Z, X_1, X_2, \ldots, X_n$ and $Y_1, \ldots, Y_n$ of Σ^* let*

$$(X_n, Y_n) \cdots (X_1, Y_1) \star Z = \{X_n^{k_n} \cdots X_1^{k_1} Z Y_1^{k_1} \cdots Y_n^{k_n} \ : \ k_i \geq 0, 1 \leq i \leq n\}$$

We make the following notational extension:

Definition 4.2.7 *For subsets $Z, X_1, X_2, \ldots, X_n$ and $Y_1, \ldots, Y_n$ of Σ^* and integers $j_1, j_2, \cdots, j_n \geq 0$ let*

$$(X_n, Y_n)^{[j_n]} \cdots (X_1, Y_1)^{[j_1]} \star Z = X_n^{j_n} \cdots X_1^{j_1} Z Y_1^{j_1} \cdots Y_n^{j_n}$$

Now that we have established the closure properties for these restricted cases, we recall an additional lemma from Ginsburg and Spanier [17] (also in Ginsburg [14, Theorem 5.3.1(a), pp. 155–157]):

Lemma 4.2.8 *$L \subseteq w_1^* w_2^*$ is a bounded context-free language if and only if L is a finite union of sets of the form $(x_m, y_m) \cdots (x_1, y_1) \star z$ where x_i is in w_1^*, y_i is in w_2^* and z is in $w_1^* w_2^*$*

Thus, for any CFL $L_1 \subseteq w_1 w_2$, we may write

$$L_1 = \bigcup_{i=1}^{N} \Big((w_1^{a_{i,m_i}}, w_2^{b_{i,m_i}}) \cdot (w_1^{a_{i,m_i-1}}, w_2^{b_{i,m_i-1}}) \cdots (w_1^{a_{i,1}}, w_2^{b_{i,1}}) \star (w_1^{a_{i,0}} w_2^{b_{i,0}}) \Big)$$

Finally, we must define analogues of the sets from Definition 3.3.4, and analogues of arithmetic progressions over n-tuples of integers:

Definition 4.2.9 *Let $L \subseteq w_1^* w_2^*$ be a bounded language. Given fixed integers N, m_i $(1 \leq i \leq N)$, and $a_{i,j}$ and $b_{i,j}$ $(1 \leq i \leq N,\ 1 \leq j \leq m_i)$ we define, for each $1 \leq i \leq N$ the set*

$$K_i(L) = \{(j_{m_i}, \ldots, j_1) \ : \ (w_1^{a_{i,m_i}}, w_2^{b_{i,m_i}})^{[j_{m_i}]} \cdots (w_1^{a_{i,1}}, w_2^{b_{i,1}})^{[j_1]} \star (w_1^{a_{i,0}} w_2^{b_{i,0}}) \in L\}$$

Definition 4.2.10 *Let $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ be n-tuples of integers. Then we call the set*

$$\{(a_1, \ldots, a_n)n + (b_1, \ldots, b_n) \ : \ n \geq 0\}$$

an n-ary arithmetic progression.

Thus, we have the following partial characterization of bounded languages with minimal context-free covers, analogous to Theorem 3.3.5:

Theorem 4.2.11 *Let $L' \subseteq L \subseteq w_1^* w_2^*$ be bounded languages. Let L be a CFL, with representation*

$$L = \bigcup_{i=1}^{N} \Big((w_1^{a_{i,m_i}}, w_2^{b_{i,m_i}}) \cdot (w_1^{a_{i,m_i-1}}, w_2^{b_{i,m_i-1}}) \cdots (w_1^{a_{i,1}}, w_2^{b_{i,1}}) \star (w_1^{a_{i,0}} w_2^{b_{i,0}}) \Big)$$

Then L' has minimal context-free cover L if and only if for all $1 \leq i \leq N$, the set $K_i(L')$ intersects every m_i-ary arithmetic progression.

Proof: First, let $L = \bigcup_{i=1}^{N} L_i$ where

$$L_i = (w_1^{a_{i,m_i}}, w_2^{b_{i,m_i}}) \cdot (w_1^{a_{i,m_i-1}}, w_2^{b_{i,m_i-1}}) \cdots (w_1^{a_{i,1}}, w_2^{b_{i,1}}) \star (w_1^{a_{i,0}} w_2^{b_{i,0}}) \qquad (4.2.1)$$

And let $L'_i = L' \cap L_i$. By Theorem 4.2.1 and Theorem 4.2.5 we have that L_i is a minimal context-free cover of L'_i if and only if L is a minimal context-free cover of L'. Note that since $L' = \bigcup_{i=1}^{N} L'_i$, it suffices to show the result for L_i, by Theorem 4.2.4.

First, assume that for L'_i, $K = K_i(L'_i)$ does not intersect every m_i-ary arithmetic progression. Let $\mathbf{c} = (c_1, \ldots, c_{m_i})$ and $\mathbf{d} = (d_1, \ldots, d_{m_i})$ be m_i-tuples of integers with $\mathbf{c} \neq \mathbf{0}$ such that $\mathbf{c}r + \mathbf{d} \notin K$ for all $r \geq 0$. Define D as the set

$$\{(w_1^{a_{i,m_i}}, w_2^{b_{i,m_i}})^{[c_{m_i} \cdot r + d_{m_i}]} \cdots (w_1^{a_{i,1}}, w_2^{b_{i,1}})^{[c_1 \cdot r + d_1]} \star (w_1^{a_{i,0}} w_2^{b_{i,0}}) \ : \ r \geq 0\}$$

Clearly D is infinite since $\mathbf{c} \neq \mathbf{0}$, and it is given by the CFG $G = (\{S\}, \Sigma, P, S)$ with productions P given by

$$S \rightarrow w_1^{A_1} S w_2^{B_1} | w_1^{A_2} w_2^{B_2}$$

where A_1, B_1, A_2, B_2 are integer constants defined by

$$\begin{aligned} A_1 &= \sum_{k=1}^{m_i} a_{i,k} \cdot c_k \\ A_2 &= a_{i,0} + \sum_{k=1}^{m_i} a_{i,k} \cdot d_k \\ B_1 &= \sum_{k=1}^{m_i} b_{i,k} \cdot c_k \\ B_2 &= b_{i,0} + \sum_{k=1}^{m_i} b_{i,k} \cdot d_k \end{aligned}$$

Thus, D is context-free. As D is also bounded by $w_1^* w_2^*$, we know that $L_i - D$ is also context-free. Clearly $D \subseteq L_i$ and $D \cap L_i' = \emptyset$, by definition of D. Since D contains only strings which do not contribute to K, and strings which contribute to K lie in L_i', we must have that $D \cap L_i' = \emptyset$.

Thus, define L_i'' as $L_i'' = L_i - D$. Then we have L_i'' is context-free, $L_i' \subseteq L_i'' \subseteq L_i$ and $L_i - L_i'' = D$ is infinite. Thus L_i is not a minimal context- free cover of L_i.

To prove the converse, we assume that L_i' is not minimally covered by L_i. Then there must exist a CFL L_i'' such that $L_i' \subseteq L_i'' \subseteq L_i$ is infinite. Recall from (4.2.1) that

$$L_i = (w_1^{a_{i,m_i}}, w_2^{b_{i,m_i}}) \cdot (w_1^{a_{i,m_i-1}}, w_2^{b_{i,m_i-1}}) \cdots (w_1^{a_{i,1}}, w_2^{b_{i,1}}) \star (w_1^{a_{i,0}} w_2^{b_{i,0}})$$

Now, from Theorem 4.2.2, we have that $L_i - L_i''$ is a context-free language. Thus, write

$$L_i - L_i'' = \bigcup_{\ell=1}^{M} \Big((w_1^{c_{\ell,m_\ell}}, w_2^{d_{\ell,m_\ell}}) \cdot (w_1^{c_{\ell,m_\ell-1}}, w_2^{d_{\ell,m_\ell-1}}) \cdots (w_1^{c_{\ell,1}}, w_2^{d_{\ell,1}}) \star (w_1^{c_{\ell,0}} w_2^{d_{\ell,0}}) \Big)$$

Since $L_i - L_i''$ is infinite, we must have that there is an ℓ, $1 \leq \ell \leq M$ such that

$$(w_1^{c_{\ell,m_\ell}}, w_2^{d_{\ell,m_\ell}}) \cdot (w_1^{c_{\ell,m_\ell-1}}, w_2^{d_{\ell,m_\ell-1}}) \cdots (w_1^{c_{\ell,1}}, w_2^{d_{\ell,1}}) \star (w_1^{c_{\ell,0}} w_2^{d_{\ell,0}}) \qquad (4.2.2)$$

is infinite. Further, there must be a k, $1 \leq k \leq m_\ell$ such that one of $d_{\ell,k}$ or $c_{\ell,k}$ is non-zero, since if this were not so (4.2.2) would not be infinite. We can thus choose arbitrary $j_1, j_2, \cdots, j_{k-1}, j_{k+1}, \cdots, j_{m_\ell}$ such that if we define the set D to be

$$D = \{(w_1^{c_{\ell,m_\ell}}, w_2^{d_{\ell,m_\ell}})^{[j_{m_\ell}]} \cdot (w_1^{c_{\ell,m_\ell-1}}, w_2^{d_{\ell,m_\ell-1}})^{[j_{m_\ell-1}]} \cdots (w_1^{c_{\ell,k+1}}, w_2^{d_{\ell,k+1}})^{[j_{k+1}]}$$
$$(w_1^{c_{\ell,k}}, w_2^{d_{\ell,k}})^{[m]} (w_1^{c_{\ell,k-1}}, w_2^{d_{\ell,k-1}})^{[j_{k-1}]} \cdots (w_1^{c_{\ell,1}}, w_2^{d_{\ell,1}})^{[j_1]} \star (w_1^{c_{\ell,0}} w_2^{d_{\ell,0}}) \ : \ m \geq 0\},$$

then D is infinite. Since $D \subseteq L_i$, each string of D is a member of L_i. We want to express each member of D in terms of $a_{i,j}$ and $b_{i,j}$.

Now, consider (4.2.2). It is contained in L_i by definition, and so each string in it is also in L_i. By definition of the $\star$ operator, $w_1^{c_{\ell,0}} w_2^{d_{\ell,0}} \in L_i$ so that we can find integers $J_{0,r}$, $1 \leq r \leq m_i$, so that

$$w_1^{c_{\ell,0}} w_2^{d_{\ell,0}} = (w_1^{a_{i,m_i}}, w_2^{b_{i,m_i}})^{[J_{0,m_i}]} \cdots (w_1^{a_{i,1}}, w_2^{b_{i,1}})^{[J_{0,1}]} \star (w_1^{a_{i,0}} w_2^{b_{i,0}})$$

Similarly, $w_1^{c_{\ell,s}} w_1^{c_{\ell,0}} w_2^{d_{\ell,0}} w_2^{c_{\ell,s}}$ is a member of L_i for each $1 \leq s \leq m_\ell$. So we can find $J_{s,r}$, $1 \leq r \leq m_i$ for each $1 \leq s \leq m_\ell$ such that

$$w_1^{c_{\ell,s}} w_1^{c_{\ell,0}} w_2^{d_{\ell,0}} w_2^{c_{\ell,s}} = (w_1^{a_{i,m_i}}, w_2^{b_{i,m_i}})^{[J_{0,m_i}+J_{s,m_i}]} \cdots (w_1^{a_{i,1}}, w_2^{b_{i,1}})^{[J_{0,1}+J_{s,1}]} \star (w_1^{a_{i,0}} w_2^{b_{i,0}})$$

Define the following constants:

$$J_j = J_{0,j} + \sum_{s=1, s \neq k}^{m_\ell} j_s \cdot J_{s,j}$$

It is then an easy observation that

$$\begin{aligned} &(w_1^{a_{i,m_i}}, w_2^{b_{i,m_i}})^{[J_{m_i}+m \cdot J_{k,m_i}]} \cdots (w_1^{a_{i,1}}, w_2^{b_{i,1}})^{[J_1+m \cdot J_{k,1}]} \star (w_1^{a_{i,0}} w_2^{b_{i,0}}) \\ = \quad & (w_1^{c_{\ell,m_\ell}}, w_2^{d_{\ell,m_\ell}})^{[j_{m_\ell}]} \cdot (w_1^{c_{\ell,m_\ell-1}}, w_2^{d_{\ell,m_\ell-1}})^{[j_{m_\ell-1}]} \cdots (w_1^{c_{\ell,k+1}}, w_2^{d_{\ell,k+1}})^{[j_{k+1}]} \cdot \\ & (w_1^{c_{\ell,k}}, w_2^{d_{\ell,k}})^{[m]} (w_1^{c_{\ell,k-1}}, w_2^{d_{\ell,k-1}})^{[j_{k-1}]} \cdots (w_1^{c_{\ell,1}}, w_2^{d_{\ell,1}})^{[j_1]} \star (w_1^{c_{\ell,0}} w_2^{d_{\ell,0}}) \end{aligned}$$

Then we have expressed each string in D in terms of $a_{i,j}$ and $b_{i,j}$.

Finally, each string $t \in D$ is a member of $L_i - L_i''$ and thus, since $L_i' \subseteq L_i''$, $t \notin L_i'$. Thus, if we let let $\mathbf{u} = (J_{m_i}, \ldots, J_1)$ and $\mathbf{v} = (J_{k,m_i}, \ldots, J_{k,1})$, by definition of $K_i(L_i')$,

$$\mathbf{v}r + \mathbf{u} \notin K_i(L_i')$$

for all $r \geq 0$. Thus $K_i(L_i')$ does not intersect every m_i-ary arithmetic progression, completing the proof. □

As a simple example of Theorem 4.2.11, consider:

Example 4.2.12: Let $L = \{a^i b^i \ : \ i \neq n + n! \ \forall n \geq 0\}$. Then if $L_1 = \{a^n b^n \ : \ n \geq 0\}$, L has minimal context-free cover L_1. This follows since $L_1 = (a, b) \star (\epsilon)$ and $K(L) = \{i \ : \ i \neq n + n! \ \forall n \geq 0\}$, which intersects every arithmetic progression.

More generally, let $k_1, k_2, k_3, c_1, c_2, c_3$ be arbitrary integer constants. Let

$$L = \{a^{k_1 n + k_2 m + k_3} b^{c_1 n + c_2 m + c_3} \ : \ n + m \text{ is not prime}\}.$$

Then if $L_1 = (a^{k_1}, b^{c_1})(a^{k_2}, b^{c_2}) \star (a^{k_3} b^{c_3})$,

$$K(L) = \{(n, m) \ : \ n + m \text{ is not prime}\}.$$

Certainly $K(L)$ intersects every binary arithmetic progression. To see this, let $(a_0, a_1)n + (b_0, b_1)$ be an arbitrary binary arithmetic progression. Choosing $n = (b_0 + b_1)$ gives a value of the progression which is not prime, as it is divisible by $b_0 + b_1$, and thus lies in $K(L)$. So L has minimal context-free cover L_1. □

Thus, Theorem 4.2.11 completely characterizes languages bounded by two words which possess a minimal context-free cover. The theorem relies on the closure of context-free languages under intersection when one of the languages is bounded by two words. This closure under intersection is certainly not true when we consider even three words, thus while a complete characterization of bounded languages possessing a minimal context-free cover might be possible, Theorem 4.2.11 is the best we can hope for using the above techniques.

4.3 Palindromic images of MRCs

In this section, we devise a method for constructing languages with minimal context-free covers from an arbitrary language with a minimal regular cover. For any language $L \subseteq \Sigma^*$, define

$$\mathbf{PAL}(L) = \{xx^R \ : \ x \in L\}$$

$$\mathbf{PAL}^{-1}(L) = \{x \in \Sigma^* \ : \ xx^R \in L\}$$

Note that if R is a regular language, $\mathbf{PAL}(R)$ is a CFL. This suggests the following theorem:

Theorem 4.3.1 *Let $L, R \subseteq \Sigma^*$. If R is a minimal regular cover for L, then $\mathbf{PAL}(R)$ is a minimal context-free cover for $\mathbf{PAL}(L)$.*

Proof: First, note that $\mathbf{PAL}(L) \subseteq \mathbf{PAL}(R)$, since if $xx^R \in \mathbf{PAL}(L)$, $x \in L \subseteq R$ so $xx^R \in \mathbf{PAL}(R)$. Let L' be a CFL such that

$$\mathbf{PAL}(L) \subseteq L' \subseteq \mathbf{PAL}(R).$$

Consider $\mathbf{PAL}^{-1}(L')$. If $x \in L$ then $xx^R \in \mathbf{PAL}(L) \subseteq L'$. Thus, $x \in \mathbf{PAL}^{-1}(L')$. Similarly, if $x \in \mathbf{PAL}^{-1}(L')$ then $xx^R \in L' \subseteq \mathbf{PAL}(R)$. Thus $\mathbf{PAL}^{-1}(L') \subseteq R$, giving

$$L \subseteq \mathbf{PAL}^{-1}(L') \subseteq R.$$

If we can argue that $\mathbf{PAL}^{-1}(L')$ is a regular set, we are finished, since then we have found a regular language $\mathbf{PAL}^{-1}(L')$ between L and R, and thus since R is a minimal regular cover for L, we must have $R - \mathbf{PAL}^{-1}(L')$ is finite. But

$$\mathbf{PAL}(R - \mathbf{PAL}^{-1}(L')) = \mathbf{PAL}(R) - L'.$$

since $\mathbf{PAL}(\mathbf{PAL}^{-1}(L')) = L'$ by our choice of L'. Thus, we have that $\mathbf{PAL}(R) - L'$ is finite. So, $\mathbf{PAL}(R)$ is a minimal context-free cover for $\mathbf{PAL}(L)$, subject to the regularity of $\mathbf{PAL}^{-1}(L')$. □

It remains to show that $\mathbf{PAL}^{-1}(L')$ is a regular set. From Horváth *et al.* [25], we say a language L is palindromic if each $x \in L$ is a palindrome. We naturally extend this to say a language is even palindromic if each $x \in L$ is an even palindrome. Consider the following theorem of Horváth *et al.* [25]:

Theorem 4.3.2 *A CFL $L \subseteq \Sigma^*$ is palindromic if and only if it is of the form*

$$L = \bigcup_{a \in \Sigma \cup \{\epsilon\}} \{waw^R \ : \ w \in L_a\}$$

for some regular languages L_a.

This proves the result, since for our $L' \subseteq \mathbf{PAL}(R)$, L' is palindromic, and since all strings in L' are of even length, we must have that $\mathbf{PAL}^{-1}(L')$ is regular.

Now, let $L_1 \subseteq L$ with L an even palindromic CFL. By the results of Horváth *et al.* [25], $\mathbf{PAL}^{-1}(L)$ is regular. Further, $\mathbf{PAL}^{-1}(L_1) \subseteq \mathbf{PAL}^{-1}(L)$. Then we have the following:

Theorem 4.3.3 *Let $L_1, L \subseteq \Sigma^*$, with L an even palindromic CFL and L a minimal context-free cover for L_1. Then $\mathbf{PAL}^{-1}(L)$ is a minimal regular cover for $\mathbf{PAL}^{-1}(L_1)$.*

Proof: Let R be a regular language such that $\mathbf{PAL}^{-1}(L_1) \subseteq R \subseteq \mathbf{PAL}^{-1}(L)$. Then it is an easy observation that $L_1 \subseteq \mathbf{PAL}(R) \subseteq L$ and since $\mathbf{PAL}(R)$ is a CFL, we must have that

$$L - \mathbf{PAL}(R)$$

is finite. Then, since

$$\mathbf{PAL}^{-1}(L - \mathbf{PAL}(R)) = \mathbf{PAL}^{-1}(L) - R$$

we have that $\mathbf{PAL}^{-1}(L) - R$ must be finite. Thus, $\mathbf{PAL}^{-1}(L)$ is a minimal regular cover for $\mathbf{PAL}(L_1)$. □

Thus we see that L has minimal regular cover R if and only if $\mathbf{PAL}(L)$ has minimal context-free cover $\mathbf{PAL}(R)$. We can now make the following observation:

Corollary 4.3.4 *Let $u, v, w, x, y \in \Sigma^*$ be fixed words with $v, x \neq \epsilon$. Let*

$$L = \{uv^iwx^iyy^R(x^R)^iw^R(v^R)^iu^R \;:\; i \geq 0\}$$

Then L has a minimal context-free cover with respect to finite languages if and only if L is context-free.

Proof: Let $L' = \{uv^iwx^iy \ : \ i \geq 0\}$. Then $L = \mathbf{PAL}(L')$ is context-free if and only if L' is regular, by Horváth *et al.* [25]. By Theorem 3.6.5, we know that L' is regular if and only if L' has a minimal regular cover, namely L' itself. Then by Theorem 4.3.1, L' has minimal regular cover L' if and only if L has minimal context-free cover L. □

4.4 Letter-equivalency and MCFCs

We recall the Parikh mapping and letter-equivalency. Our presentation is based on Harrison [22]. Let $\Sigma = \{a_1, a_2, \ldots, a_n\}$ be our fixed alphabet. Then define $\psi : \Sigma^* \to \mathbb{N}^n$ by

$$\psi(w) = (|w|_{a_1}, |w|_{a_2}, \ldots, |w|_{a_n})$$

We naturally extend this to languages as $\psi(L) = \{\psi(w) \ : \ w \in L\}$.

A subset of $\mathbb{N}^n$ is said to be linear if it is of the form

$$\{\alpha_0 + n_1\alpha_1 + n_2\alpha_2 + \cdots + n_m\alpha_m \ : \ n_i \geq 0 \text{ for } 1 \leq i \leq m\}$$

for elements $\alpha_0, \alpha_1, \ldots, \alpha_m \in \mathbb{N}^n$. A set is said to be semi-linear if it as finite union of linear sets. We say that L has semi-linear image if $\psi(L)$ is semi-linear.

We say that two languages $L_1, L_2 \subseteq \Sigma^*$ are letter-equivalent if $\psi(L_1) = \psi(L_2)$. Also note that if $L_1 \subseteq L_2$ then $\psi(L_1) \subseteq \psi(L_2)$.

The following results are well known (eg, see Harrison [22, pp. 226–230]):

Lemma 4.4.1 *A language L has a semilinear image if and only if L is letter equivalent to a regular set.*

The following is known as Parikh's Theorem [40]:

Theorem 4.4.2 *Every context-free language has semilinear image.*

Thus, every context-free language is letter equivalent to a regular language. Now we are ready for the section's main theorem:

Theorem 4.4.3 *Let $L, L_1 \subseteq w_1^* w_2^*$ be languages, with L_1 a minimal context-free cover of L. Then L is letter equivalent to a language L' which has a minimal regular cover R_1. Further, R_1 is letter equivalent to L_1.*

Proof: Let

$$L_1 = \bigcup_{i=1}^{N} \left((w_1^{a_{i,m_i}}, w_2^{b_{i,m_i}}) \cdot (w_1^{a_{i,m_i-1}}, w_2^{b_{i,m_i-1}}) \cdots (w_1^{a_{i,1}}, w_2^{b_{i,1}}) \star (w_1^{a_{i,0}} w_2^{b_{i,0}}) \right),$$

where $\star$ is given in Definition 4.2.6. Then $K_i(L)$ intersects every m_i-ary arithmetic progression, by Theorem 4.2.11. Now, define R_1 to be the language

$$R_1 = \bigcup_{i=1}^{N} \left((w_1^{a_{i,m_i}} w_2^{b_{i,m_i}})^* \cdot (w_1^{a_{i,m_i-1}} w_2^{b_{i,m_i-1}})^* \cdots (w_1^{a_{i,1}} w_2^{b_{i,1}})^* (w_1^{a_{i,0}} w_2^{b_{i,0}}) \right).$$

R_1 is regular and it is easily observed that R_1 is letter equivalent to L_1. Define $L' = \{x \in R_1 \ : \ \psi(x) \in \psi(L)\}$. We claim that L' has minimal regular cover R_1. L' is letter equivalent to L by construction.

We use Theorem 3.3.5 to show that L' has minimal regular cover R_1. To do so, let

$$R_{1,i} = (w_1^{a_{i,m_i}} w_2^{b_{i,m_i}})^* \cdot (w_1^{a_{i,m_i-1}} w_2^{b_{i,m_i-1}})^* \cdots (w_1^{a_{i,1}} w_2^{b_{i,1}})^* (w_1^{a_{i,0}} w_2^{b_{i,0}}).$$

Each is certainly regular and bounded by itself. If we define $L'_i = R_{1,i} \cap L'$ then by Theorems 3.1.4 and 3.1.1, it suffices to show that L'_i has minimal regular cover $R_{1,i}$.

Now, let k be an arbitrary integer satisfying $1 \leq k \leq m_i$. Consider

$$I = I_k(L', j_1, \ldots, j_{k-1}, j_{k+1}, \ldots, j_{m_i})$$

for some integers $j_1, \ldots, j_{k-1}, j_{k+1}, \ldots, j_{m_1}$. Assume there is an arithmetic progression $am + b$ such that $am + b \notin I$ for all $m \geq 0$. We now construct an m_i-ary arithmetic progression which does not intersect $K_i(L)$. This will contradict our assumption. The m_i-ary arithmetic progression is

$$\begin{aligned}&(j_1, \ldots, j_{k-1}, am + b, j_{k+1}, \ldots, j_{m_i})\\ &\quad = (j_1, \ldots, j_{k-1}, b, j_{k+1}, \ldots, j_{m_i}) + m(0, \ldots, 0, a, 0, \ldots, 0)\end{aligned}$$

Assume that $(j_1, \ldots, j_{k-1}, am + b, j_{k+1}, \ldots, j_{m_i}) \in K_i(L)$ for some $m \geq 0$. Then

$$\begin{aligned}&(w_1^{a_{i,m_i}}, w_2^{b_{i,m_i}})^{[j_{m_i}]} \cdot (w_1^{a_{i,m_i-1}}, w_2^{b_{i,m_i-1}})^{[j_{m_i}]} \cdots (w_1^{a_{i,k+1}}, w_2^{b_{i,k+1}})^{[j_{k+1}]} \cdot\\ &\quad (w_1^{a_{i,k}}, w_2^{b_{i,k}})^{[am+b]} \cdot (w_1^{a_{i,k-1}}, w_2^{b_{i,k-1}})^{[j_{k-1}]} \cdots (w_1^{a_{i,1}}, w_2^{b_{i,1}})^{[j_1]} \star (w_1^{a_{i,0}} w_2^{b_{i,0}}) \in L.\end{aligned}$$

Thus, the string

$$(w_1^{a_{i,m_i}} w_2^{b_{i,m_i}})^{j_{m_i}} \cdot (w_1^{a_{i,m_i-1}} w_2^{b_{i,m_i-1}})^{j_{m_i}} \cdots (w_1^{a_{i,k+1}} w_2^{b_{i,k+1}})^{j_{k+1}} \cdot$$
$$(w_1^{a_{i,k}} w_2^{b_{i,k}})^{am+b} \cdot (w_1^{a_{i,k-1}} w_2^{b_{i,k-1}})^{j_{k-1}} \cdots (w_1^{a_{i,1}} w_2^{b_{i,1}})^{j_1} (w_1^{a_{i,0}} w_2^{b_{i,0}}) \in L'$$

is in L', by definition of L', and $am + b \in I$ by definition of I. □

Theorem 4.4.3 leads us to conjecture the following:

Conjecture 4.4.4 *Let $L, L_1 \subseteq \Sigma^*$ be languages, with L_1 context-free and $L \subseteq L_1$. If L has minimal context-free cover L_1 then L is letter equivalent to a language L' such that*

(a). L' has a minimal regular cover R and

(b). R is letter equivalent to L_1.

The converse of the conjecture is not true, as we see in the next lemma.

Lemma 4.4.5 *Let $L = \{0^n 1^{2n} 0^n \ : \ n \geq 0\}$ Then L does not have a minimal context-free cover, but L is letter equivalent to a regular language.*

Proof: Corollary 4.3.4 gives that L has a minimal context-free cover if and only if L is context-free, by setting $u = w = y = \epsilon$, $v = 0$, and $x = 1$. But L is easily seen to not be context-free, as $L = h^{-1}(a^n b^n c^n) \cap (0^+(1^2)^+0^+)$, where $h(a) = 0$, $h(b) = 11$ and $h(c) = 0$. Thus L has no minimal context-free cover.

However, L is semilinear, as $\psi(L) = \{n(2, 2) \ : \ n \geq 0\}$. Thus by Theorem 4.4.2, L is letter equivalent to a regular language, for example $(0^2 1^2)^*$, which certainly has minimal regular cover itself. □

In trying to prove Conjecture 4.4.4, Theorems 4.3.1 and 4.3.3 will not give us a counter-example, since languages with minimal context-free cover produced by these theorems will always be letter-equivalent to a homomorphic image of the language that were used to create them.

4.5 Minimal co-CF Covers

From the results in Section 4.3, we can characterize a class of languages which possess a minimal co-context-free cover, via the following:

Corollary 4.5.1 *Let $u, v, w, x, y \in \Sigma^*$ be fixed words with $v, x \neq \epsilon$. Let*

$$L = \{uv^iwx^iyy^R(x^R)^iw^R(v^R)^iu^R \;:\; i \geq 0\}$$

*Then L has a infinite context-free subset if and only if L is a CFL. Further, if L is not a CFL, $\overline{L}$ has minimal co-**CF** cover Σ^*.*

Proof: We first show that L has an infinite context-free subset if and only if L is itself context-free. Obviously if L is a CFL then L itself is a context-free subset of L.

To prove the converse, assume $L' \subseteq L$ is a context-free subset of L. Then by the results of Horváth *et al.* [25], we must have that $\mathbf{PAL}^{-1}(L')$ is a regular language, since $L' \subseteq L$ is even palindromic. But note that

$$\mathbf{PAL}^{-1}(L') \subseteq \mathbf{PAL}^{-1}(L) = \{uv^iwx^iy \;:\; i \geq 0\}$$

Further, $\mathbf{PAL}^{-1}(L')$ is infinite since L' is. Thus by 3.6.4, we must have $\mathbf{PAL}^{-1}(L)$ is regular. We conclude that L is a CFL.

Thus, assume that L is not a CFL, and so has no infinite context-free subset. Then for all context-free languages L_0 such that $\overline{L} \subseteq \overline{L_0} \subseteq \Sigma^*$, observe that $L_0 \subseteq L$ and thus L_0 is finite. This shows that Σ^* is a minimal co-context-free cover for $\overline{L}$.

□

This is obviously analogous to Theorem 3.6.6 for minimal regular covers. However, we have a result pertaining to co-**CF** covers due to the fact that the CFLs are not closed under complement.

Chapter 5

Open Problems

We conclude by examining the open problems relating to minimal regular covers and the related work examined in this thesis.

5.1 Open Problems for MRCs

Despite the work done in Sections 3.1 and 3.2, the closure properties of minimal covers under inverse ϵ-free regular substitution remains open. Given a substitution $\varphi : \Sigma \to 2^{\Delta^*}$, we recall that there are two possible definitions of inverse substitution: either

$$\varphi^{-1}(L) = \{x \in \Sigma^* \ : \ \varphi(x) \subseteq L\}$$

or

$$\varphi^{[-1]}(L) = \{x \in \Sigma^* \ : \ \varphi(x) \cap L \neq \emptyset\}.$$

In Section 3.1, an example was given which showed that, under the first definition, there is a language L with minimal regular cover R and a ϵ-free substitution φ with $\varphi^{-1}(L)$ not minimally covered by the regular language $\varphi^{-1}(R)$. However, it may still be that $\varphi^{-1}(L)$ has another minimal regular cover which is a proper subset of $\varphi^{-1}(R)$. Thus, we have the following open problems:

Open Problem 5.1.1 *Given a language L with a minimal regular cover, and an ϵ-free regular substitution φ, does $\varphi^{-1}(L)$ have a minimal regular cover?*

With respect to the second definition of inverse substitution, we only have a partial solution to the problem in the unary case, as Theorem 3.1.16 gives that for any unary language L with minimal regular cover 0^*, $\varphi^{[-1]}(L)$ has minimal regular cover Σ^*. Thus, for the unary case, we have the following:

Open Problem 5.1.2 *Given a unary language $L \subseteq 0^*$ with a minimal regular cover $R \subseteq 0^*$, $R \neq 0^*$ and a ϵ-free regular substitution $\varphi : \Sigma^* \to 2^{0^*}$, does $\varphi^{[-1]}(L)$ have minimal regular cover $\varphi^{[-1]}(R)$?*

For non-unary languages, the same problem is also open:

Open Problem 5.1.3 *Given a language $L \subseteq \Delta^*$ with a minimal regular cover R and a ϵ-free regular substitution $\varphi : \Sigma^* \to 2^{\Delta^*}$, does $\varphi^{[-1]}(L)$ have minimal regular cover $\varphi^{[-1]}(R)$?*

5.1.1 Bi-reg-immune sets and MRCs

This thesis examined in great detail the case of a language L having a minimal regular cover L. However, in Section 2.2.1, we saw the definition of bi-immunity for a class of languages. Considering minimal covers to be an extension of immunity, we have proposed the following definition:

Definition 5.1.4 *A language L has bi-minimal $\mathcal{C}$-cover (L_1, L_2) with respect to $\mathcal{D}$ if L_1 is a minimal $\mathcal{C}$-cover of L with respect to $\mathcal{D}$ and L_2 is a minimal $\mathcal{C}$-cover of $\overline{L}$ with respect to $\mathcal{D}$.*

Again, the case where $\mathcal{C} = \mathbf{REG}$ and $\mathcal{D} = \mathbf{FIN}$ would be of primary interest. However, there is much more consideration to be given to this topic than that devoted to it in Section 3.7.

For instance, in the unary case, we defined a class of context-sensitive languages which were bi-**REG**-immune. Since the context-sensitive languages are the smallest class of languages which properly contain the regular languages in the unary case, this gives the best result possible in the unary case. We may ask the same questions of larger alphabets:

Open Problem 5.1.5 *Does there exist an alphabet Σ with $|\Sigma| > 1$ such that there is a context-free language $L \subseteq \Sigma^*$ with bi-minimal regular cover (R_1, R_2)?*

We know by Theorem 2.3.5 that if such a context-free language exists, it is necessarily nondeterministic.

5.2 Open decision problems for MRCs

In Section 3.4, we proved that the following problems were undecidable:

- Given an arbitrary CFL L, does L have a minimal regular cover?

- Given an arbitrary CFL L, does $\overline{L}$ have a minimal cover?

- Given an arbitrary linear CFG G, does $L(G)$ have a minimal regular cover?

- Given an arbitrary linear CFG G, does $\overline{L(G)}$ have a minimal regular cover?

Clearly, the decision problem "Given a regular language R, does R have a minimal regular cover?" is trivially true. Thus, we are left with the DCFLs as the only natural class of languages to which we do not know the decidability status of the 'minimal regular cover possession' problem.

This remains open: Is the problem "Given a DCFL L, does L have a minimal regular cover?" decidable? Since the DCFLs are closed under complement, this is equivalent to asking "Given a DCFL L, does $\overline{L}$ have a minimal regular cover?". Greibach's theorem, which proved the problem undecidable for CFLs, is not applicable to DCFLs, as "$= \Sigma^*$" is decidable for DCFLs. In fact, we have several decision problems which are decidable for DCFLs [24, Thm 10.6, p. 246-247]:

- Given a DCFL L and a regular language R, does $R = L$ [15]?

- Given a DCFL L, is L regular [45, 47]?

- Given a DCFL L and a regular language R, is $L \subseteq R$ [15]?

However, the following are undecidable [24, Thm 10.7, p. 247], as proved by Ginsburg and Greibach [15]:

- Given DCFLs L_1 and L_2, is $L_1 \cap L_2$ a DCFL?

- Given DCFLs L_1 and L_2, is $L_1 \cap L_2 = \emptyset$?

- Given DCFLs L_1 and L_2, is $L_1 \subseteq L_2$?

Since the decision problem of possession of a minimal cover only involves one DCFL (namely the L in question), it would seem reasonable to conjecture the following:

Conjecture 5.2.1 *Let M be an arbitrary DPDA. Then the decision problem 'does $L(M)$ have a minimal regular cover?' is decidable.*

5.3 Open Problems for MCFCs

We have considered several problems for minimal context-free covers. Unfortunately, several remain open. Notably, we have the following open problems

Open Problem 5.3.1 *Are minimal context-free covers unique up to finite modification?*

Open Problem 5.3.2 *What are the closure properties of languages which possess a minimal context-free cover?*

We also have our conjecture relating minimal context-free covers and letter equivalency:

Conjecture 5.3.3 *Let $L, L_1 \subseteq \Sigma^*$ be languages, with L_1 context-free and $L \subseteq L_1$. If L has minimal context-free cover L_1 then L is letter equivalent to a language L' such that*

(a). L' has a minimal regular cover R and

(b). R is letter equivalent to L_1.

Note that if this conjecture is true, if a language did have distinct context-free covers L_1 and L_2, the symmetric difference of $\psi(L_1)$ and $\psi(L_2)$ would be finite.

5.4 Open problems concerning density

Also, recall the problem of Bucher [9], which we saw in Section 2.4: given context-free languages L_1 and L_3 with $L_1 \subseteq L_3$ and $L_3 - L_1$ infinite, must there exist a context-free language L_2 such that $L_1 \subseteq L_2 \subseteq L_3$ and both $L_3 - L_2$ and $L_2 - L_1$ infinite? Apparently, there has been no progress on this problem since 1980.

Bibliography

[1] J.-P. Allouche and J. Shallit. *Automatic Sequences.* In preparation, 2000.

[2] J. L. Balcázar, J. Díaz, and J. Gabarró. *Structural Complexity I.* Springer-Verlag, 1995.

[3] J. L. Balcázar and U. Schöning. Bi-immune set for complexity classes. *Math. Systems Theory*, 18:1–10, 1985.

[4] L. Berman and J. Hartmanis. On isomorphisms and density of NP and other complete sets. *SIAM J. Comput.*, 6:305–332, 1977.

[5] J. Berstel. Every iterated morphism yields a co-CFL. *Inform. Process. Lett.*, 22(1):7–9, 1986.

[6] F. Blanchard and P. Kòrka. Language complexity of rotations and Sturmian sequences. *Theor. Comput. Sci.*, 1998:179–193, 1998.

[7] M. Blum. A machine-independent theory of the complexity of recursive functions. *J. Assoc. Comput. Mach.*, 14:322–336, 1967.

[8] M. Blum. On effective procedures for speeding up algorithms. *J. Assoc. Comput. Mach.*, 18(2):290–305, 1971.

[9] W. Bucher. A density problem for context-free languages. *Bull. European Assoc. Theor. Comput. Sci*, 10:53, 1980.

[10] C. Câmpeanu, N. Sântean, and S. Yu. Minimal cover-automata for finite languages. In J.-M. Champarnaud, D. Maurel, and D. Ziadi, editors, *WIA '98*, volume 1660 of *Lecture Notes in Computer Science*, pages 43–56. Springer-Verlag, 1998.

[11] J. C. E. Dekker. Two notes on recursively enumerable sets. *Proc. Amer. Math. Soc.*, 4:495–501, 1953.

[12] N. J. Fine and H. S. Wilf. Uniqueness theorems for periodic functions. *Proc. Amer. Math. Soc.*, 16:109–114, 1965.

[13] P. Flajolet and J. M. Steyaert. On sets having only hard subsets. In J. Loeckx, editor, *Automata, Languages and Programming*, pages 446–457, 1974.

[14] S. Ginsburg. *The Mathematical Theory of Context Free Languages*. McGraw-Hill, 1966.

[15] S. Ginsburg and S. Greibach. Deterministic context free languages. *Information and Control*, 9:620–648, 1966.

[16] S. Ginsburg and S. A. Greibach. Abstract families of languages. In S. Ginsburg, S. A. Greibach, and J. Hopcroft, editors, *Studies in Abstract Families of Languages*, pages 1–32. American Mathematical Society, 1969.

[17] S. Ginsburg and E. H. Spanier. Mappings of languages by two-tape devices. *J. Assoc. Comput. Mach.*, 12:423–434, 1965.

[18] S. Ginsburg and E. H. Spanier. Bounded regular sets. *Proc. Amer. Math. Soc.*, 17:1043–1049, 1966.

[19] J. N. Gray and M. A. Harrison. On the covering and reduction problems for context-free grammars. *J. Assoc. Comput. Mach.*, 19:675–698, 1972.

[20] S. Greibach. A note on the undecidable properties of formal languages. *Math. Systems Theory*, 2:1–6, 1968.

[21] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers.* Oxford University Press, 1989.

[22] M. A. Harrison. *Introduction to Formal Language theory.* Addison-Wesley, 1978.

[23] J. Hartmanis. *Feasible Computations and Provable Complexity Properties.* SIAM, 1978.

[24] J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation.* Addison-Wesley, 1979.

[25] S. Horváth, J. Karhumäki, and J. Kleijn. Results concerning palindromicity. *J. Inf. Process. Cybern. EIK*, pages 441–451, 1987.

[26] H.B. Hunt and D. J. Rosenkrantz. Computational parallels between the regular and context-free languages. *SIAM J. Comput.*, 7:99–114, 1978.

[27] H.B. Hunt and D.J. Rosenkrantz. On equivalence and containment problems for formal languages. *J. Assoc. Comput. Mach.*, 24(3):387–396, 1977.

[28] H.B. Hunt, D.J. Rosenkrantz, and T.G. Szymanski. On the equivalence, containment, and covering problems for the regular and context-free languages. *J. Comput. System Sci.*, 12:222–268, 1976.

[29] N. Immerman. Nondeterministic space is closed under complementation. *SIAM J. Comput.*, 17:935–938, 1988.

[30] K. Ko and D. Moore. Completeness, approximation and density. *SIAM J. Comput.*, 10:787–796, 1981.

[31] M. Latteux and G. Thierrin. Semidiscrete context-free languages. *Internat. J. Comput. Math.*, 14(1):3–18, 1983.

[32] M. Latteux and G. Thierrin. On bounded context-free languages. *Elektron. Informationsverarb. Kybernet.*, 20(1):3–8, 1984.

[33] M. Lothaire. *Algebraic Combinatorics on Words.* In preparation, 2000.

[34] J. H. Lutz. Category and measure in complexity classes. *SIAM J. Comput*, 19:1100–1131, 1990.

[35] J. H. Lutz. Almost everywhere high nonuniform complexity. *J. Comput. System Sci.*, 44:220–258, 1992.

[36] N. Lynch. On reducibility to complex or sparse sets. *J. Assoc. Comput. Mach*, 22:341–345, 1975.

[37] R.C. Lyndon and M. P. Schützenberger. The equation $a^M = b^N c^P$ in a free group. *Michigan Math.*, 9:289–298, 1962.

[38] E. Mayordomo. Almost every set in exponential time is P-bi-immune. *Theor. Comput. Sci.*, 136:487–506, 1994.

[39] C. H. Papadimitriou. *Computational Complexity.* Addison-Wesley, 1994.

[40] R. J. Parikh. On context-free languages. *J. Assoc. Comput. Mach.*, 13:570–581, 1966.

[41] A. Paun, N. Santean, and S. Yu. An $O(n^2)$ algorithm for constructing minimal cover automata for finite languages. In M. Daley, M. G. Eramian, and S. Yu, editors, *Pre-Proceedings, CIAA 2000*, pages 233–241, 2000.

[42] E. L. Post. Recursively enumerable sets of integers and their decision problems. *Bull. Amer. Math. Soc.*, 50:284–316, 1944.

[43] H. Rogers. *Theory of Recursive Functions and Effective Computability.* McGraw-Hill, 1967.

[44] A. Salomaa. *Formal Languages.* Academic Press, 1973.

[45] R.E. Stearns. A regularity test for pushdown machines. *Information and Control*, 11:323–340, 1967.

[46] R. Szelepcsényi. The method of forced enumeration for nondeterministic automata. *Acta Informatica*, 26:279–284, 1988.

[47] L. Valiant. Regularity and related problems for deterministic pushdown automata. *J. Assoc. Comput. Mach.*, 22(1):1–10, 1975.